

Durham Research Online

Deposited in DRO:

22 October 2020

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Cuesta-Lazaro, Carolina and Li, Baojiu and Eggemeier, Alexander and Zarrouk, Pauline and Baugh, Carlton M. and Nishimichi, Takahiro and Takada, Masahiro (2020) 'Towards a non-Gaussian model of redshift space distortions.', *Monthly notices of the Royal Astronomical Society.*, 498 (1). pp. 1175-1193.

Further information on publisher's website:

<https://doi.org/10.1093/mnras/staa2249>

Publisher's copyright statement:

This article has been accepted for publication in *Monthly notices of the Royal Astronomical Society*. ©: 2020 The Author(s). Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Towards a non-Gaussian model of redshift space distortions

Carolina Cuesta-Lazaro,^{1,2★} Baojiu Li^{1b}, Alexander Eggemeier,¹ Pauline Zarrouk,¹ Carlton M. Baugh,^{1,2} Takahiro Nishimichi^{3,4} and Masahiro Takada⁴

¹*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

²*Institute for Data Science, Durham University, South Road, Durham DH1 3LE, UK*

³*Center for Gravitational Physics, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan*

⁴*Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*

Accepted 2020 July 2. Received 2020 June 22; in original form 2020 February 3

ABSTRACT

To understand the nature of the accelerated expansion of the Universe, we need to combine constraints on the expansion rate and growth of structure. The growth rate is usually extracted from 3D galaxy maps by exploiting the effects of peculiar motions on galaxy clustering. However, theoretical models of the probability distribution function (PDF) of galaxy pairwise peculiar velocities are not accurate enough on small scales to reduce the error on theoretical predictions to the level required to match the precision expected for measurements from future surveys. Here, we improve the modelling of the pairwise velocity distribution by using the Skew-T PDF, which has non-zero skewness and kurtosis. Our model accurately reproduces the redshift space multipoles (monopole, quadrupole, and hexadecapole) predicted by N -body simulations, above scales of about $10 h^{-1} \text{Mpc}$. We illustrate how a Taylor expansion of the streaming model can reveal the contributions of the different moments to the clustering multipoles, which are independent of the shape of the velocity PDF. The Taylor expansion explains why the Gaussian streaming model works well in predicting the first two redshift space multipoles, although the velocity PDF is non-Gaussian even on large scales. Indeed, any PDF with the correct first two moments would produce precise results for the monopole down to scales of about $10 h^{-1} \text{Mpc}$, and for the quadrupole down to about $30 h^{-1} \text{Mpc}$. An accurate model for the hexadecapole needs to include higher order moments.

Key words: dark energy – large-scale structure of Universe – cosmology: miscellaneous – cosmology: theory.

1 INTRODUCTION

By combining different experiments, such as the supernova standard candles (Perlmutter et al. 1997; Riess et al. 1998, 2007), and the cosmic microwave background (CMB) temperature anisotropies (Planck Collaboration VI 2018), astronomers have inferred that the expansion of the Universe is accelerating. If general relativity is the correct theory of gravity, concordance with the aforementioned experiments requires that about 95 per cent of the energy content of the universe is invisible to us, since it does not exhibit electromagnetic interactions. Most of this energy is postulated to be in the form of a cosmological constant, Λ , which drives the observed accelerated expansion of the Universe. Its nature is, however, unknown.

A compelling theoretical explanation attributes the cosmological constant to the energy of empty space. None the less, its estimated value (Carroll 2001) has produced the largest discrepancy between theoretical predictions and observations ever encountered in physics. Therefore, despite the success of the standard model of cosmology [known as Lambda cold dark matter (Λ CDM)] in explaining observations, we still do not understand the nature of some of its constituents, which has led cosmologists to look for evidence beyond Λ CDM.

Most of the extensions to the standard model add new degrees of freedom to either the energy content of the Universe or to the way

space-time geometry reacts to it, by modifying general relativity (Clifton et al. 2012; Joyce, Lombriser & Schmidt 2016). The most popular example of extra degrees of freedom in the energy content is dynamical dark energy that allows for temporal variation of the dark energy equation of state, driven by a scalar field. The second class of models involves modifications to the Einstein–Hilbert action, and is commonly denoted as modified gravity. The modified action adds new terms to the Poisson equation, which are identified with the appearance of an additional fifth force. However, the boundary between the two categories is not always well defined.

To distinguish between these different scenarios observationally, tests of the background expansion history are not sufficient. Viable gravity theories can be tuned to reproduce the observed evolution of the scale factor with time, and therefore are indistinguishable on the background level. However, by including information about the rate at which cosmic structures grow, we can detect modifications to gravity. The growth of cosmic structure is the outcome of a competition between the expansion of the Universe and the gravitational pull, generated by inhomogeneities. If there is an additional fifth force, but the expansion is compatible with that observed, the rate at which structures in the Universe grow will be modified.

To estimate the large-scale growth of structure in the Universe, we look at the statistics of 3D galaxy maps. These maps contain the angular position of galaxies on the sky, together with their redshifts. Assuming that galaxies are at rest, as the photons they emit travel towards us through an expanding Universe, their wavelengths stretch

★ E-mail: carolina.cuesta-lazaro@durham.ac.uk

accordingly. Therefore, we observe the redshifted light of distant galaxies. We can translate this redshift into a comoving distance by introducing the Hubble factor, $H(z)$,

$$r(z) = \int_0^z \frac{dz'}{H(z')}, \quad (1)$$

where $r(z)$ is the comoving distance to the galaxy, and we have used the natural unit where the speed of light $c = 1$.

However, galaxies also move due to the gravitational pull generated by the inhomogeneous distribution of matter around them. If a source that emits light moves, the wavelength of the light gets further redshifted due to the Doppler effect. If we ignore this effect, then we would infer the wrong distance, s , given by

$$s = \mathbf{r} + \frac{\mathbf{v}(\mathbf{r})\hat{z}}{\mathcal{H}}\hat{z}, \quad (2)$$

instead of the real position of the galaxy, \mathbf{r} , where $\mathbf{v}(\mathbf{r})$ is the peculiar velocity of the galaxy, $\mathcal{H} = aH(a)$ the comoving Hubble factor, and the inferred distance, s , is the redshift space distance. Note that we have assumed the observer is far away from the sources, and therefore the line-of-sight direction can be fixed to a particular direction, which we arbitrarily set as the \hat{z} -axis. This approximation, known as the plane-parallel approximation, has so far given results that lie within the statistical error bars of current surveys (Samushia, Percival & Raccanelli 2012). However, this approximation will need to be used more carefully in the analysis of upcoming surveys.

The translation between redshift and distance is in reality then more complex than equation (1), since we need to disentangle the combination of the galaxy's position and its peculiar velocity along the line of sight. However, this complication ends up being beneficial because peculiar velocities are generated by the gravitational pull of the inhomogeneous matter distribution, and therefore allow us to extract information on the latter. In particular, as mentioned above, it is possible to detect the existence or constrain the strength of fifth forces by studying the growth of structure inferred from the statistics of the peculiar velocity field (see e.g. Guzzo et al. 2008; Bose & Koyama 2017; Hernández-Aguayo et al. 2019).

To extract the growth rate, we measure the effect of peculiar velocities on clustering known as redshift space distortions (RSDs). The amount of clustering is quantified as the probability of finding a pair of galaxies at a given separation, compared to the probability for a random distribution of galaxies. This statistic is known as the two-point correlation function, defined as

$$\xi^R(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle, \quad (3)$$

which, due to statistical isotropy and homogeneity, only depends on pair separation, r . The two-point correlation function is a useful statistic because it summarizes all the statistical information contained in a Gaussian random field. Even though the evolved density field is non-Gaussian, it can still be used to constrain the cosmological parameters. However, the higher order correlation functions of the evolved density field are non-zero, and hence contain additional independent information over and above that encoded within the variance (see e.g. Song, Taruya & Oka 2015; Gagrani & Samushia 2017, in the context of redshift space clustering).

Due to the peculiar motions of galaxies, we observe redshift space positions, s , instead of the real space positions, \mathbf{r} , and thus we can only measure

$$\xi^S(s_\perp, s_\parallel) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{s}) \rangle, \quad (4)$$

which depends on both the pair separation, s , and its inclination with respect to the line-of-sight direction. Throughout we denote

separations perpendicular and parallel to the line of sight as s_\perp and s_\parallel , respectively.

The so-called redshift space correlation function, $\xi^S(s_\perp, s_\parallel)$, is a combination of both real space clustering, $\xi^R(r)$, and the probability of finding a pair of galaxies with a given relative velocity along the line of sight, also denoted as pairwise velocity distribution, as we will show in Section 2 using the streaming model of RSD (see e.g. Fisher 1995; Scoccimarro 2004). Since clustering in redshift space is affected by relative peculiar motions, it contains information about the growth of structure.

In this paper, we focus on improving the accuracy of models of the redshift space correlation function, to improve our estimates of the growth rate. The main hurdle that has to be overcome is the non-linear evolution of the density and velocity fields produced by non-linearities in the continuity and Euler equations that drive gravitational collapse. As we will see, this is particularly relevant to describe the mapping of pairs from real to redshift space, which is necessary to model the two-point correlation function in equation (4). The development of this mapping is our focus here.

State-of-the-art constraints on the growth factor are found measuring the two-point correlation function in redshift space (e.g. Satpathy et al. 2017 for galaxies from BOSS; Zarrouk et al. 2018 for eBOSS quasars), which have reported growth factors consistent with general relativity. Satpathy et al. (2017) used measurements of the two-point correlation function down to separations of $25 h^{-1}\text{Mpc}$, beyond which theoretical predictions introduce larger systematic errors than the statistical errors of the measurement itself, thus biasing the estimate of the growth factor. For future surveys, the expected statistical errors will be significantly smaller (e.g. Huterer et al. 2015) and so we will need more accurate theoretical predictions down to small scales than those used in the analysis of current surveys, to improve constraints on the growth rate and to avoid making catastrophic errors of interpretation (Jennings, Baugh & Pascoli 2011).

The goal of this paper is threefold. First, we introduce an extension to the simplest streaming model that assumes Gaussian relative motions, which improves the accuracy of theoretical predictions for the clustering multipoles. Secondly, we show a comparison of state-of-the-art models for the streaming model ingredients with high-resolution N -body simulations. Finally, we analyse the effect of the different velocity moments on the clustering multipoles, and assess how accurate their theoretical predictions need to be for an RSD model that is at least as accurate as the measurements from future surveys.

This paper is structured as follows. In Section 2, we summarize the theoretical framework used to model the RSD effect on the two-point correlation function, the streaming model, and we present the Skew-T (hereafter ST) distribution as a model for the pairwise velocity distribution that includes non-Gaussian features. In Section 3, we compare the accuracy of the ST model with a simple Gaussian distribution, using N -body simulations. In Section 4, we study a Taylor expansion of the streaming model integrand, which allows us to examine how the different higher order moments affect the multipoles. In Section 5, we analyse the sensitivity of the multipoles to the different real space ingredients of the streaming model. Finally, in Section 6 we summarize our main results and draw our conclusions.

2 THE STREAMING MODEL OF REDSHIFT SPACE DISTORTIONS

The streaming model provides the mapping from the real-space two-point correlation function to the observed anisotropic two-point

correlation function in redshift space. Since objects viewed in redshift space are the same as those in real space, but have been moved to different positions, we can relate their density contrasts by imposing mass conservation

$$(1 + \delta^S(s)) d^3s = (1 + \delta^R(r)) d^3r, \quad (5)$$

where superscript S denotes redshift space, and R real space. This expression can be further manipulated (Scoccimarro 2004) to obtain a relation between real and redshift space clustering

$$1 + \xi^S(s_\perp, s_\parallel) = \int_{-\infty}^{\infty} dr_\parallel (1 + \xi^R(r)) \mathcal{P}(v_\parallel = s_\parallel - r_\parallel | r), \quad (6)$$

where $r^2 = r_\parallel^2 + r_\perp^2$, $s_\perp = r_\perp$, $\mathcal{P}(v_\parallel | r)$ is the pairwise velocity distribution, and $v_\parallel = v_{\parallel,1} - v_{\parallel,2}$, is the line-of-sight relative velocity of the pair of tracers. In our convention, v_\parallel is defined as negative (positive) if the pairs are approaching (receding from) each other. Equation (6) is known as the streaming model (Fisher 1995), which simply states that the probability of finding a pair of objects at a distance s in redshift space is given by the sum over all possible real space distances, r , that would make us infer the redshift space position. While the streaming model is one way to move forward, fully Eulerian perturbation theory treatments based on the same expression can also accurately describe redshift space clustering (Taruya, Nishimichi & Saito 2010).

The plus one terms in equation (6) ensure that given a universe with randomly placed galaxies, if the pairwise velocity distribution is dependent on the pair separation, then we would still observe clustering in redshift space induced by the coherent velocity field. If, however, the pairwise velocity distribution does not depend on pair separation, the plus one terms on both sides in equation (6) cancel out.

Note that equation (6) is exact, and the only approximation we have made so far is the plane-parallel approximation to select a particular line of sight. None the less, the apparent simplicity of the streaming model may be deceptive, as the complexity of the gravitational dynamics is hidden in the shape of $\mathcal{P}(v_\parallel | r)$ and its dependence on pair separation. Broadly speaking, on small scales within dark matter haloes, virial motions produce a large velocity dispersion that reduces the amount of clustering along the line of sight; the size of this effect increases with halo mass. On larger scales, galaxies in-falling into larger structures shift the mean velocity to negative values, producing a change in the opposite sense to those on small scales, that increases the inferred clustering along the line of sight (Kaiser 1987).

It has been known for a long time (Scoccimarro 2004) that this scenario is further complicated by the non-Gaussian nature of the pairwise velocity distribution, which is evident from its non-zero skewness and kurtosis. There is no Gaussian limit for pairwise velocities on large scales, since velocity differences cancel out long-range contributions and leave only the local, non-linear component of the velocity at the two different locations. Here, we focus on extending the streaming model to include these non-Gaussian features, as predicted by N -body simulations.

Throughout, we will use the relation between the full three-dimensional pairwise velocity and its line-of-sight projection. The line-of-sight pairwise velocity distribution can be obtained by integrating the full distribution $\mathcal{P}(v_r, v_t | r)$, where the radial velocity, v_r , and the transverse velocity, v_t , are defined as the velocity components parallel and transverse to the pair separation vector, respectively. Due to statistical isotropy, we only need to select one component from the two-dimensional transverse velocity. For ease of computation, we chose the one which will contribute to the line-of-sight projection,

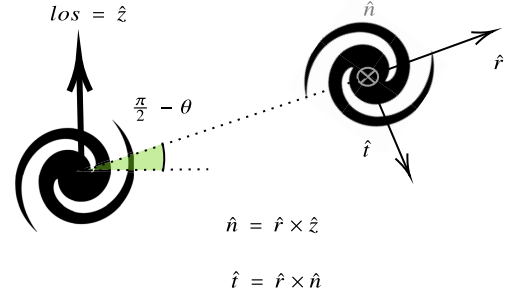


Figure 1. Decomposition of the 3D distance vector into a radial component along the pair distance, \hat{r} , a normal component, \hat{n} , which is perpendicular to both the line-of-sight direction and the pair separation vector, and a transverse component, \hat{t} , which completes the basis formed by the radial and normal vectors. After projecting the distance vector on to the line of sight, only the radial and transverse components will give a non-zero contribution.

i.e. the one in the plane spanned by the galaxy pair and the observer (see Fig. 1). Thus,

$$v_\parallel = v_r \cos \theta + v_t \sin \theta, \quad (7)$$

where θ is the angle between the pair separation vector and the line of sight, $\theta = \tan^{-1}(r_\perp/r_\parallel)$. Therefore,

$$\mathcal{P}(v_\parallel | r_\perp, r_\parallel) = \int \frac{dv_r}{\sin \theta} \mathcal{P}\left(v_r, v_t = \frac{v_\parallel - v_r \cos \theta}{\sin \theta} \middle| r\right). \quad (8)$$

The relations between the moments of the two distributions are given by

$$c_n(r_\perp, r_\parallel) = \sum_{k=0}^n \binom{n}{k} \mu^k (1 - \mu^2)^{\frac{n-k}{2}} c_{k,n-k}(r), \quad (9)$$

where c_n denotes the n -th central moment of the line-of-sight projected distribution, $\mathcal{P}(v_\parallel | r_\perp, r_\parallel)$, and $c_{k,n-k}$ the k -th radial moment, $(n-k)$ -th transverse moment of $\mathcal{P}(v_r, v_t | r)$, and $\mu = \cos \theta$. The n -th moment about the origin is denoted as m_n .

2.1 The Gaussian streaming model

The commonly used model for the redshift space correlation function is known as the Gaussian streaming model (GSM; Fisher 1995; Reid & White 2011). The radial and transverse components of the pairwise velocity are taken to be independently Gaussian distributed. The line-of-sight projection can therefore be written as

$$\mathcal{P}_G(v_\parallel | r) = \frac{1}{\sqrt{2\pi\sigma_{12}^2(r)}} \exp\left[-\frac{(v_\parallel - v_{12}(r))^2}{2\sigma_{12}^2(r)}\right], \quad (10)$$

where $v_{12}(r)$, denoted as $m_1(r)$ in our notation, and $\sigma_{12}(r)$, equivalent to $\sqrt{c_2(r)}$, are the projections of the radial and transverse moments on to the line of sight, and are both dependent on the pair separation.

As explained in the previous section, a Gaussian distribution does not accurately describe the pairwise velocity distribution for an evolved matter distribution even for large pair separations. However, this simplified assumption gives an accurate description of the clustering of dark matter haloes on scales larger than $30 h^{-1} \text{Mpc}$ (Reid & White 2011; Wang, Reid & White 2014). Later on, we shall illustrate how the accuracy of this model stems from the integral in equation (6) over the pairwise velocity distribution, which on large scales only receives contributions from the lowest order pairwise velocity moments.

None the less, an accurate model on smaller scales requires non-vanishing higher order moments, mainly the skewness and kurtosis.

Different approaches have been taken towards such a model in the literature. On one hand Uhlemann, Kopp & Haugg (2015) performed an Edgeworth expansion around a Gaussian distribution to add skewness, and found improvements with respect to the GSM on scales smaller than $30 h^{-1} \text{Mpc}$. We provide a more in-depth discussion of this model on the following section. On the other hand, a number of authors (e.g. Sheth 1996; Tinker 2007; Bianchi, Chiesa & Guzzo 2015; Bianchi, Percival & Bel 2016; Kuruvilla & Porciani 2018) have all used mixtures of normal or quasi-normal distributions to model a skewed and heavy-tailed distribution. The first approach by Sheth (1996) modelled the motion of the halo centre of mass using a Maxwellian distribution which is then weighted by the Press–Schechter mass function. Tinker (2007) developed a similar approach using the halo model (Cooray & Sheth 2002), but assuming that, at fixed environmental density around the halo pair, the pairwise velocity distribution of haloes is Gaussian. The skewness is then developed by weighting these Gaussian distributions with the probability of finding a given density. The parameters of the model are calibrated using N -body simulations.

Further developments were introduced by Bianchi et al. (2015), who replaced the mixing distribution described above by another Gaussian, which assumes that the mean and standard deviations of the ‘local’ Gaussian distributions are themselves jointly distributed according to a bivariate Gaussian. This model, however, cannot generate distributions that are sufficiently skewed to explain the halo pairwise velocity distribution. This limitation was later overcome by performing an Edgeworth expansion on the local distributions, which added skewness to the Gaussian distribution (Bianchi et al. 2016).

A more recent study by Kuruvilla & Porciani (2018) used a generalized hyperbolic distribution (GHD) to model the pairwise velocity distribution of N -body simulations. In this case, the relation between the parameters of the distribution and velocity moments as a function of pair separation is not given, and the model requires five free parameters with a two-dimensional dependency on the pair separation vector.

2.2 The Edgeworth streaming model

The Edgeworth Streaming Model introduced by Uhlemann et al. (2015) is one of the simplest extensions to the GSM. The authors used an Edgeworth expansion of the velocity PDF to extend the validity of the GSM towards smaller scales. The Edgeworth expansion is an asymptotic series expansion to a probability density function, which implies that there is no guarantee of convergence when more terms are added to the expansion. See Sellentin, Jaffe & Heavens (2017) for an interesting discussion on the Edgeworth expansion and its applications to cosmology.

Expanding the line of sight velocity PDF around a Gaussian distribution one obtains, to first order,

$$\mathcal{P}_E(v_{\parallel}|\mathbf{r}) = \frac{1}{\sqrt{2\pi\sigma_{12}^2(\mathbf{r})}} \exp\left[-\frac{(v_{\parallel} - v_{12}(\mathbf{r}))^2}{2\sigma_{12}^2(\mathbf{r})}\right] \times \left(1 + \frac{\Lambda_{12}}{6\sigma_{12}^3} H_3\left(\frac{v_{\parallel} - v_{12}}{\sigma_{12}}\right)\right), \quad (11)$$

where Λ_{12} is the third-order cumulant of the velocity PDF projected on to the line-of-sight direction, and H_3 the third-order probabilists’ Hermite polynomial, $H_3(x) = x^3 - 3x$.

In the next section, we present a flexible model that we believe is simpler than the ones mentioned above and achieves similar or better levels of accuracy.

2.3 The Skewed Student-t (ST) streaming model

A study of the cluster-galaxy cross-correlation by Zu & Weinberg (2013) found that the skewed Student-t distribution (ST; Azzalini & Capitanio 2003) gives an accurate description of the cluster-galaxy pairwise velocity statistics predicted by simulations. The main advantage of using this distribution to model RSD is that its parameters can be written as functions of the four lowest order moments. Here, we use the ST distribution to model the redshift space clustering of galaxy or halo pairs on all scales.

In recent years, there has been increasing interest in such flexible probability density functions that can accommodate different degrees of skewness and kurtosis. More specifically, a successful approach proposed by Azzalini & Capitanio (2003), found that a skewed, multivariate, distribution can be generated by combining a symmetric density function with a cumulative distribution function as follows:

$$f(x) = 2 f_0(x) G(w(x)), \quad x \in \mathbb{R}^d, \quad (12)$$

where $f_0(x)$ is a symmetric PDF defined in \mathbb{R}^d , G is a one-dimensional cumulative distribution function, whose derivative satisfies $G'(x) = G'(-x)$, and w is a real-valued odd function in \mathbb{R} .

Since we are interested in a distribution that displays both skewness and extended tails, the symmetric function f_0 can be chosen to be a ST distribution, hereafter referred to as the t -distribution, which in one dimension is given by

$$f_0(x) = t_1(x - x_c|w, \nu) := \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} w \Gamma(\frac{\nu}{2})} \left(1 + \frac{1}{\nu} \left(\frac{x - x_c}{w}\right)^2\right)^{-\frac{\nu+1}{2}}. \quad (13)$$

The t -distribution is characterised by three parameters: the location x_c , the shape parameter, w , and the number of degrees of freedom, ν . The latter controls the decay of probability in the tails, and therefore allows us to describe distributions with varying degrees of kurtosis.

The skewed multivariate distribution which originates from the t -distribution by using equation (12) is known as the ST distribution, hereafter ST. Its density function for a one-dimensional random variable, x , is,

$$f_{\text{ST}}(x|x_c, w, \alpha, \nu) := \frac{2}{w} t_1(x - x_c|1, \nu) T_1\left[\alpha \frac{(x - x_c)}{w} \left(\frac{\nu + 1}{\nu + \left(\frac{x - x_c}{w}\right)^2}\right)^{1/2}; \nu + 1\right], \quad (14)$$

where t_1 is the one-dimensional t -distribution defined by equation (13), and T_1 is the one-dimensional cumulative t -distribution with $\nu + 1$ degrees of freedom. The ST distribution has an extra skewness parameter, α , compared to the t -distribution.

Note that the dependence of the distribution parameters on the pair separation vector, \mathbf{r} , has been omitted for clarity. The relation between these parameters and the four lowest order moments of the ST distribution can be found in Appendix A.

3 COMPARISON WITH N-BODY SIMULATIONS

In this section, we assess the performance of the different RSD models by comparing them to a set of dark matter only Λ CDM simulations.

3.1 Simulations

We use the Dark Quest (Nishimichi et al. 2019) set of simulations, which consists of 15 independent realizations of the density

fluctuations in a cosmological volume, adopting the best-fitting cosmological parameters given by the Planck CMB data (Planck Collaboration XIII 2016)

$$\{\omega_b, \omega_c, \Omega_{DE}, \ln(10^{10} A_s), n_s, w_{DE}\} \\ = \{0.02225, 0.1198, 0.6844, 3.094, 0.9645, -1\}, \quad (15)$$

where $\omega_b \equiv \Omega_b h^2$, and $\omega_c \equiv \Omega_c h^2$ are the physical density parameters of baryons and cold dark matter, respectively, $\Omega_{DE} = 1 - (\omega_b + \omega_c + \omega_\nu)/h^2$ is the dark energy density parameter (assuming a flat Universe and the neutrino density parameter, ω_ν corresponding to the total mass of 0.06 eV for the three neutrino species), A_s and n_s are the amplitude and tilt of the primordial curvature power spectrum normalized at 0.05 Mpc^{-1} , and w_{DE} is the equation-of-state parameter of dark the energy.

The simulations follow the evolution of 2048^3 particles in a comoving box of size $L = 2h^{-1} \text{ Gpc}$, which translates into a particle mass of $m_p = 8.158 \times 10^{10} h^{-1} M_\odot$, using the Tree-Particle Mesh code GADGET2 (Springel 2005). Halo catalogues were constructed using the publicly available ROCKSTAR halo finder (Behroozi, Wechsler & Wu 2013). Here, we focus on accurate predictions for massive central haloes, with masses above $10^{13} h^{-1} M_\odot$, and leave the predictions for galaxies to future work. In all figures below, we show the mean simulation measurements over the 15 independent realizations of the cosmological volume, with errorbars representing one standard deviation of the mean measurements. All results are shown for the $z = 0$ snapshots.

3.2 The ingredients of the streaming model

The streaming model, equation (6), takes as input both the pairwise velocity distribution and the real space two-point correlation function. In this subsection, we will show the measurements of both ingredients from the simulations, together with their theoretical predictions for the given cosmological parameters.

Using the halo catalogues from the simulations, we measure the pairwise velocity distribution in bins of $0.5 h^{-1} \text{ Mpc}$ size (note that the velocities are rescaled by \mathcal{H} so that they have the unit of length). As mentioned above, in our convention the pairwise velocity is defined as negative (positive) when the members of the pair are approaching (receding from) each other. We show the measured pairwise distribution from the simulations, for a few selected cases of $(r_\perp, r_{||})$, in Fig. 2. The figure shows increasing $r_{||}$ values from left to right, and increasing r_\perp from top to bottom.

The black dots in Fig. 2 show the measured pairwise velocity of dark matter haloes, while the lines give the ST (red) and Gaussian (blue) distributions obtained by applying two different methods to find the best-fitting parameters, which will be described in Section 3.3. It is evident that in all cases the Gaussian distributions are a poorer fit to the simulation measurements than the ST distributions. In particular, by comparing the symbols with the blue curves, we note that for all pair separations there is a significant kurtosis in the simulation data which a Gaussian distribution fails to capture.

In the cases of $r_{||} = 5.25 h^{-1} \text{ Mpc}$ and $r_\perp = 0.75$ or $5.25 h^{-1} \text{ Mpc}$, the pairwise velocity distributions are also very strongly skewed towards negative $v_{||}$, which is because such close halo pairs are more likely to be found in high-density regions where haloes approach each other ($v_{||} < 0$), than in void regions where haloes tend to move apart ($v_{||} > 0$). The skewness, however, decreases for much larger r_\perp (e.g. $49.75 h^{-1} \text{ Mpc}$, the bottom panel of the central column of Fig. 2) or

$r_{||}$ (the right column of Fig. 2), because the probabilities of infalling and receding halo pairs tend to be even out for large separations. On the other hand, the left column of Fig. 2 shows that for very small $r_{||}$ (e.g. $0.75 h^{-1} \text{ Mpc}$), the skewness is small again, which is because in this case the pair separation vector is nearly perpendicular to the line of sight, and $v_{||} \approx v_t$, meaning that $v_{||}$ has equal probability to be in any direction within the plane perpendicular to the pair separation vector due to statistical isotropy, that is, equal probability of $v_{||} > 0$ and $v_{||} < 0$.

In Fig. 3, we show the radial and transverse pairwise velocity distribution for the haloes at different pair separations. Note that the two components are not independent. This figure shows the same physical picture as Fig. 2. At small pair separations, but larger than halo size, the radial component has a non-zero (negative) mean, produced by tracers infalling towards larger objects. Given that the infall velocity is different in different environments and more pairs are likely to be found in high-density environments where members of a pair tend to approach each other, the radial distribution is skewed towards negative values. At larger pair separations, the radial skewness becomes smaller, but it still has heavy tails. Due to statistical isotropy, the transverse component is symmetric and has zero mean, although it also shows broader tails than a Gaussian distribution.

The moments of the different distributions are shown in Fig. 4, where the definition of the moments is mass-weighted, since the velocity field is only measured where there are tracers, by the number density of tracers at a given separation $r = |\mathbf{r}| = |\mathbf{x}_2 - \mathbf{x}_1|$,

$$m_{ij} = \frac{\langle (1 + \delta(\mathbf{x}_1))(1 + \delta(\mathbf{x}_2))v_r^i v_t^j \rangle}{\langle (1 + \delta(\mathbf{x}_1))(1 + \delta(\mathbf{x}_2)) \rangle}, \quad (16)$$

where i and j denote the order of the moments in the radial and transverse components, respectively. For instance, the radial mean is denoted as m_{10} , the second-order transverse moment as m_{02} , and the third-order cross-correlation between the radial and the squared of the transverse component as m_{12} . The central moments are analogously defined by

$$c_{ij} = \frac{\langle (1 + \delta(\mathbf{x}_1))(1 + \delta(\mathbf{x}_2))(v_r - m_{10})^i (v_t - m_{01})^j \rangle}{\langle (1 + \delta(\mathbf{x}_1))(1 + \delta(\mathbf{x}_2)) \rangle}. \quad (17)$$

Statistical isotropy in the transverse plane implies that only moments with even powers of the transverse component are non-zero. That is c_{12} for the third-order moment, and c_{22} for the fourth.

Although it is not the objective of this paper to develop the relations between the cosmological parameters and the ingredients of the streaming model (the real-space two-point correlation and the pairwise velocity moments), for completeness we show the predictions from different methods as a summary of the recent progresses in perturbation theory. This exercise will show what stage we have reached in our efforts to predict these quantities and what still needs to be done. So far, only predictions for the first two moments of the velocity field have been successfully obtained from perturbation theory:

(i) *Linear perturbation theory* – Fisher (1995) shows that the mean pairwise velocity in linear theory is determined by the correlation between the density and velocity fields, $\langle \delta v \rangle$, due to the mass-weighting factors in equation (16). The variance, however, is determined by the velocity–velocity coupling (Gorski 1988). In the simplest flavour of Eulerian perturbation theory, there are two free parameters: the linear bias and the growth factor. Higher order corrections to the mean and the variance were computed in Reid & White (2011) by expanding the continuity and Euler equations in powers of the linear density field

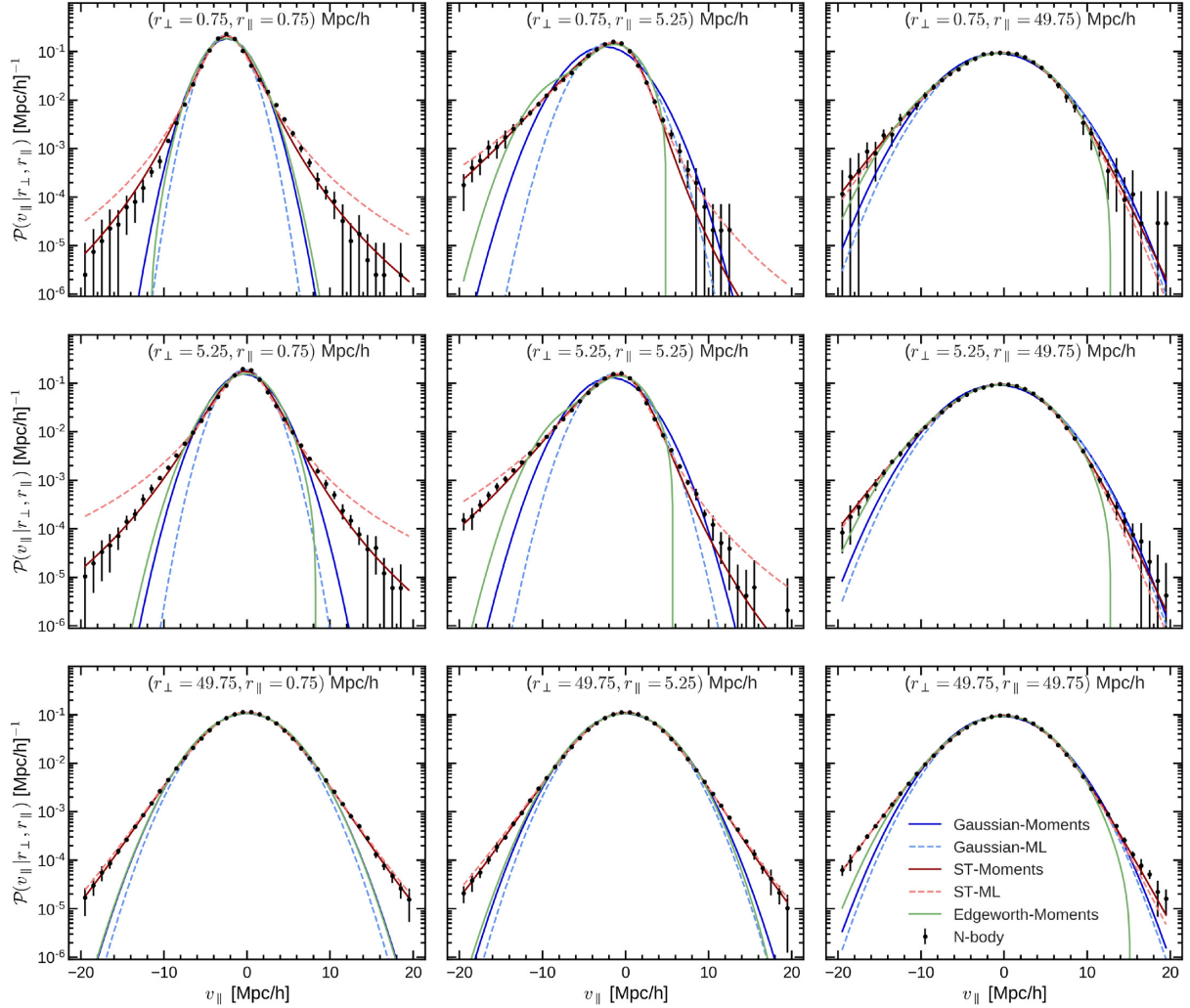


Figure 2. The pairwise line-of-sight velocity distribution for massive dark matter haloes in the simulation at $z = 0$, evaluated at different pair separations. Columns show increasing r_{\parallel} separation, whilst rows show increasing r_{\perp} . The black dots show the mean measurements from the N -body simulation and their standard deviation, whilst the solid (dashed) curves show the different models found using the method of moments (maximum likelihood) estimate. The Gaussian model is shown in blue, the Edgeworth expansion model is shown in green, and the ST model, that includes skewness and kurtosis, is shown in red.

up to fourth order. They also used a local Lagrangian prescription for the bias (Matsubara 2008), which turned out to be very important to reproduce the real space correlation function, since a local bias in Lagrangian space introduces a non-local bias in Eulerian space (Baldauf et al. 2012; Chan, Scoccimarro & Sheth 2012).

(ii) *Convolutional Lagrangian perturbation theory (CLPT)* – Wang et al. (2014) extended the formalism of Carlson, Reid & White (2013), to include predictions for the lowest order pairwise velocity moments. The Lagrangian approach formulates the problem in terms of initial positions and displacement field, where the latter fully specifies the motion of the cosmological fluid. Instead of expanding the fluid equations in terms of the linear density field, the expansion is performed on the displacement field that gives the mapping between initial Lagrangian positions and final Eulerian positions. To describe the Lagrangian bias functional, $\delta_h = F[\delta]$, the authors include three free parameters, b_1 , b_2 , and b_s , which we fit to the real-space two-point correlation function. The first two of these bias parameters, b_1 and b_2 , are the first and second derivatives of the Lagrangian bias function with respect to a long-wavelength density contrast, δ^L , whereas b_s encodes the dependence of the bias on a long-wavelength

tidal tensor. The variance of the pairwise velocities is, however, not accurately reproduced by CLPT: a constant shift needs to be added to describe the variance on linear scales. Interestingly, this constant offset is the same for both the radial and transverse components, as one would expect from the effect of virial motions. Including the growth factor, CLPT requires five parameters to describe clustering in redshift space.

(iii) *Convolutional Lagrangian effective field theory (CLEFT)* – Carrasco, Hertzberg & Senatore (2012) developed an analytical effective field theory to capture the effects of very small scales on large-scale observables. Vlah, Castorina & White (2016) used this idea, together with CLPT, to predict the lowest order velocity moments that enter the GSM. They found that predictions for the mean pairwise velocity were greatly improved compared to CLPT, especially the derivative, which ultimately controls the accuracy of the redshift space quadrupole. Moreover, it was shown that in the context of effective field theory, the constant shift in Wang et al. (2014) was identified as one of the effective parameters to describe the effect of small scales. Increased accuracy comes at the expense of requiring more free parameters, the effective

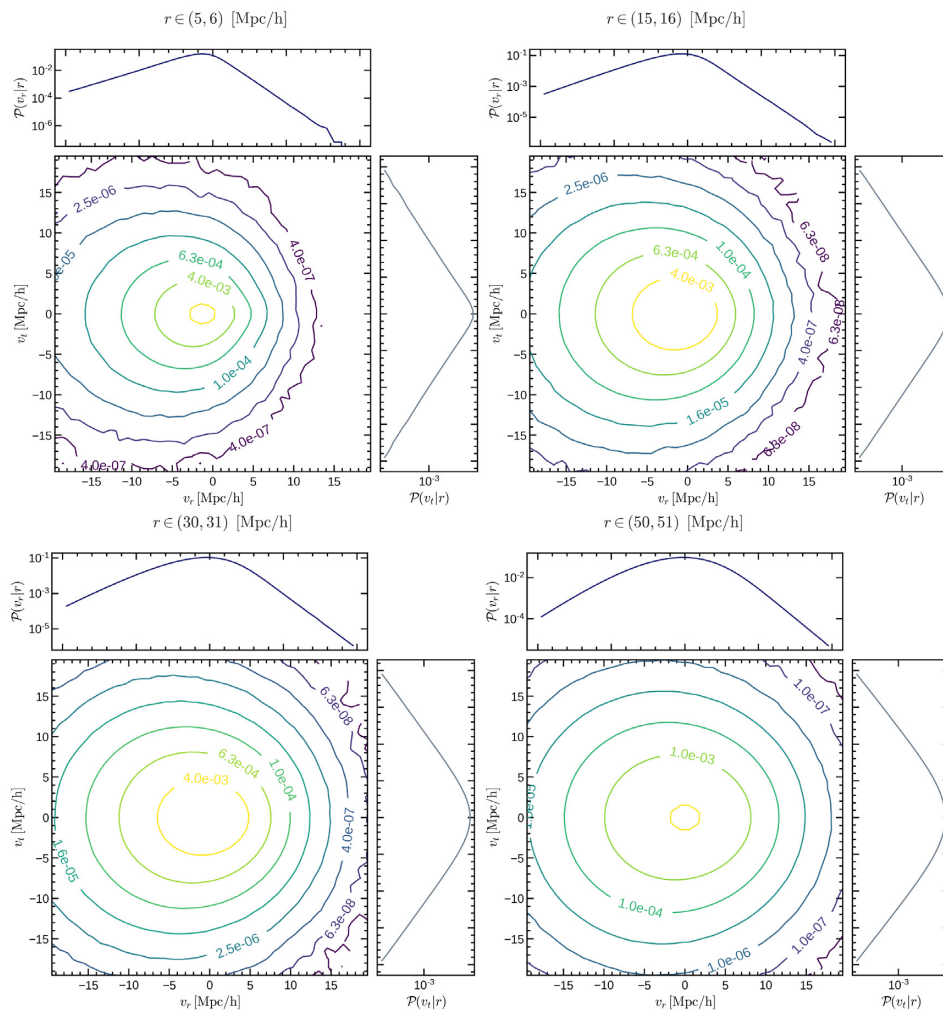


Figure 3. The mean joint probability distribution of the radial and transverse pairwise velocities of dark matter haloes measured in N -body simulations. The marginal distributions are shown on the sides. At small pair separations, infall towards larger structures produces a large skewness in the radial component, and the mean turns more negative. At large pair separations, the distributions of the two components are symmetric but still show heavy tails.

field theory counter-terms. There are two extra parameters, one for the real space correlation function and the other for the mean pairwise velocity. Therefore, the simplest CLEFT has seven parameters.

The top two panels of Fig. 4 compare the predictions for the two lowest order moments of the three different methods.¹ The symbols show measurements from simulations. In the upper panel, we show the mean of the radial pairwise velocity. The two extra EFT counter-terms extend the agreement of CLPT with N -body simulations from scales of $\sim 60 h^{-1} \text{Mpc}$ down to $\sim 20 h^{-1} \text{Mpc}$. For the radial and transverse components of the variance, shown in the second panel, the CLPT and CLEFT predictions are qualitatively similar. The reason for this is that the EFT counter-term is very close to a constant shift in the variance, which is already included in CLPT to match N -body simulation results. The moment predicted with the lowest accuracy is the radial component of the variance, where percent-level predictions are limited to scales above $40 h^{-1} \text{Mpc}$. Note that the radial component of the variance, $c_{20} = m_{20} - m_{10}^2$, has a

contribution from the mean pairwise velocity and will also be affected by errors in modelling non-linear infall.

The bottom two panels of Fig. 4 show the simulation measurements of the third- and fourth-order moments (symbols). Perturbation theory predictions for moments higher than the second have only been obtained for the third-order moment using CLPT in Uhlemann et al. (2015). However, the authors found that it fails to capture the non-Gaussian effects encoded in the skewness for scales below $100 h^{-1} \text{Mpc}$. Since third- and fourth-order moments only play a role on the accuracy of the redshift space correlation function on small scales, it is extremely difficult to produce accurate enough predictions to unlock access to the cosmological information contained on those scales. To improve these predictions, we plan to explore both effective field theory extensions to CLPT, and the use of emulators for the moments on small scales.

In Fig. 5, we also show simulation measurements (symbols) of the real-space two-point correlation function of dark matter haloes, together with the predictions using both CLPT (dashed line) and CLEFT (dash-dotted line). The CLEFT prediction is accurate over a broad range of scales – it gives percent-accuracy results on scales between 10 and $70 h^{-1} \text{Mpc}$ – at the expense of only one extra free parameter. For more details on the accuracy of the different perturbation theory models, we refer the reader to Appendix C.

¹Perturbation theory predictions have been obtained using the publicly available code `github/CLEFT.GSM`.

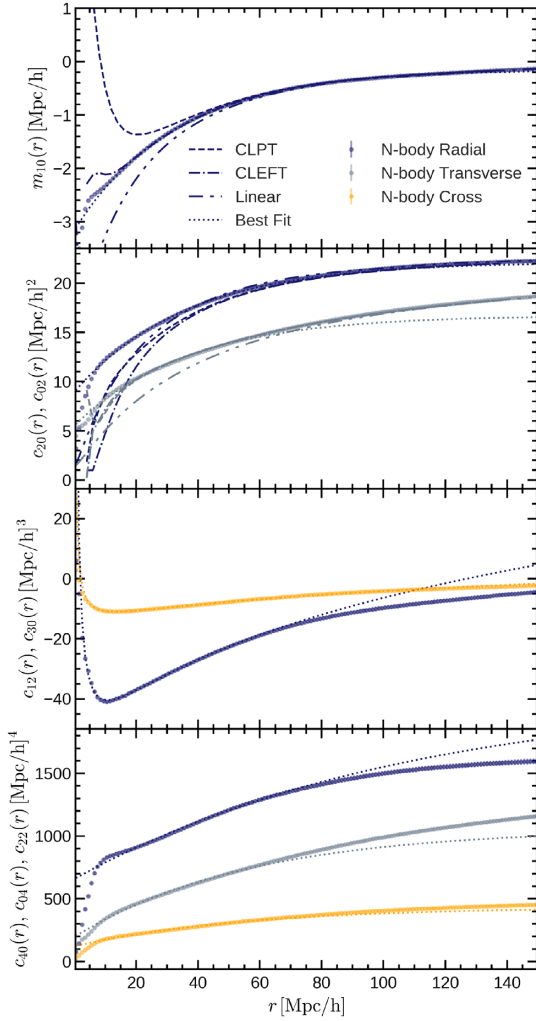


Figure 4. The four lowest order moments of the radial and transverse pairwise velocity distributions of dark matter haloes. In each panel, we show the mean measurements from the simulations, together with errorbars showing one standard deviation (note these are too small to be seen). We also show the different perturbation theory predictions for the two lowest order moments. Linear theory is shown in dot-dashed-dashed lines, CLPT in dashed lines, and CLEFT in dashed dotted lines. Finally, we show the best-fitting curves as dotted lines, which are used to show the accuracy of the Taylor expansion in Section 4.2. Note the best-fitting curves have been fitted to the moments on scales smaller than $60 h^{-1} \text{Mpc}$.

3.3 Fitting the pairwise velocity distribution

To infer the parameters of the Gaussian and ST distributions that best fit the simulation measurements, we use two different methods,

(i) *Maximum likelihood estimation*, found by maximizing the probability that the model reproduces the simulation measurements. This is equivalent to a least- χ^2 fit to the simulation measurements using the given Gaussian or ST distribution function when the errors are approximately Gaussian distributed. We refer to this method as ‘ML’ occasionally in this paper. The χ^2 we minimize is

$$\chi^2(\mathbf{r}) = \sum_{v_{\parallel}} \frac{(\mathcal{P}_{\text{measured}}(v_{\parallel}|\mathbf{r}) - \mathcal{P}_{\text{model}}(v_{\parallel}|\mathbf{r}))^2}{\mathcal{P}_{\text{model}}(v_{\parallel}|\mathbf{r})}, \quad (18)$$

where $\mathcal{P}_{\text{model}}$ is either a Gaussian or an ST distribution.

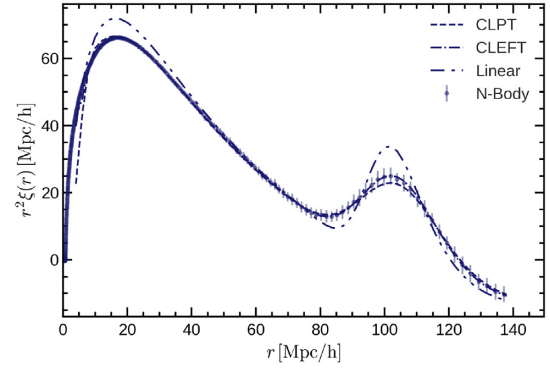


Figure 5. The real space correlation function measured from the simulations for dark matter haloes, compared with predictions from CLPT and CLEFT. Both these perturbation theory predictions use a Lagrangian prescription for the bias, and have been computed by simultaneously fitting the correlation function and the two lowest order pairwise velocity moments. For more details on the fitting see Appendix C.

(ii) *Method of moments* that uses the analytical relation between the parameters of the distribution and its lowest order moments to convert the moments estimated from the simulation measurements into distribution parameters.

If the distribution measured from the simulation and the fitted distribution are the same, both methods are equivalent. However, this is not the case when the fitted distribution is an approximation to the simulation results or when noise is present.

In Fig. 2, we show the best-fitting distributions for both the Gaussian and the ST models using these two approaches. For the Gaussian case the conversion between moments and parameters is trivial, while for the ST model we have used the relations given in Appendix A to obtain the model parameters given the four lowest order moments. In this figure, we can see that even for large pair separations the Gaussian approximation is inaccurate, where the method of moments and the maximum likelihood estimation produce slightly different results, both being poor approximations.

The ST model, however, is flexible enough to represent the varying degrees of skewness and kurtosis over a broad range of pair separations when using the method of moments. At large separations, the maximum likelihood estimate and the method of moments produce similar distributions. None the less, on small scales the tails of the distribution are mis-estimated by the maximum likelihood method.

For a more detailed comparison of the different models around the peak of the distribution see Fig. B1.

3.4 The redshift space correlation function

In this subsection, we use the Gaussian and ST models of the pairwise velocity distribution with the streaming model (equation 6) to predict redshift space clustering. We will focus on the mapping between real and redshift spaces, and show that using the more flexible ST distribution for pairwise velocity leads to more accurate predictions of the higher order redshift space multipoles than are obtained with the simpler GSM model. For this reason, we measure all the real-space quantities from the simulation, including the real-space halo two-point correlation function and the pairwise velocity distribution moments, as inputs to reproduce the redshift space clustering by using equation (6). The impact of the accuracy of the modelling of

the individual ingredients of the streaming model will be studied in a later section.

The pairwise velocity distribution has been measured in the range $0 < r_{\parallel}/[h^{-1}\text{Mpc}] < 70$ and $0 < r_{\perp}/[h^{-1}\text{Mpc}] < 50$ in bins spaced by $0.5 h^{-1}\text{Mpc}$. To perform the streaming model integration in equation (6) we have used the Simpsons rule, with a linear interpolation of the real space correlation function and the pairwise velocity distribution.

Due to the difficulty of analysing 2D plots, together with the complex covariance matrix between the different measurements for $\xi^S(s_{\perp}, s_{\parallel})$, it is common to decompose the redshift space correlation function into multipole moments using its Legendre expansion (Hamilton 1998)

$$\xi(s, \mu) = \sum_{\ell} \xi_{\ell}(\mu) L_{\ell}(\mu), \quad (19)$$

where ℓ is the order of the multipole and $L_{\ell}(\mu)$ is the Legendre polynomial at the ℓ -th order, which depends on the angular coordinate $\mu = \cos \theta$. The redshift space correlation function is symmetric in μ , so only even values of ℓ give a non-zero contribution. Inverting equation (19), we find that the multipole moments are given by

$$\xi_{\ell}(s) = \frac{2\ell + 1}{2} \int_{-1}^1 \xi(s, \mu) L_{\ell}(\mu) d\mu. \quad (20)$$

The three lowest multipoles are denoted as monopole ($\ell = 0$), quadrupole ($\ell = 2$), and hexadecapole ($\ell = 4$). Recent cosmological analyses are based mainly on the monopole and quadrupole moments; however, the cosmological information carried by the hexadecapole has also been shown to be important (Taruya, Saito & Nishimichi 2011).

We show these three multipole moments predicted by the different models, as well as the measurements from the simulations, in Fig. 6. In the lower subpanels of each panel we show the relative differences between the model predictions and the simulation results in units of the standard deviation (σ) (middle subpanel) calculated using the 15 simulation realizations each of which has a volume of $8 (h^{-1}\text{Gpc})^3$, and the relative percent error in the lowest subpanel. The yellow horizontal shaded bands represent the $\pm 1\sigma$ ranges on the multipoles.

Surprisingly, the two Gaussian distributions that we found by using the method of moments and the maximum likelihood estimate yield multipoles that can be more than five standard deviations away from each other. Furthermore, the Gaussian distribution obtained using the method of moments reproduces the three multipoles within one standard deviation for scales larger than approximately $30 h^{-1}\text{Mpc}$, although it gives a very poor fit to the pairwise velocity distribution on these scales; cf. Fig. 2.

Regarding the ST model, although the maximum likelihood ST lies closer than the Gaussian method of moments to the pairwise velocity measured in the simulation (see Fig. 2), it gives a biased result for the multipoles. On the other hand, the ST model found by the method of moments is able to reproduce the correct clustering down to scales of around $10 h^{-1}\text{Mpc}$.

The Edgeworth model does improve the predictions of the multipoles compared to the GSM, however to extend its validity to even smaller scales we need to also add fourth-order moments.

Note that, as a result of the large simulated volume, the error bars on the monopole and quadrupole on scales below $20 h^{-1}\text{Mpc}$ are extremely small, meaning that the 1σ deviations for the monopole and quadrupole (the yellow horizontal bands in the lower subpanels of Fig. 6) are within one percent of the mean measurement up to $20 h^{-1}\text{Mpc}$.

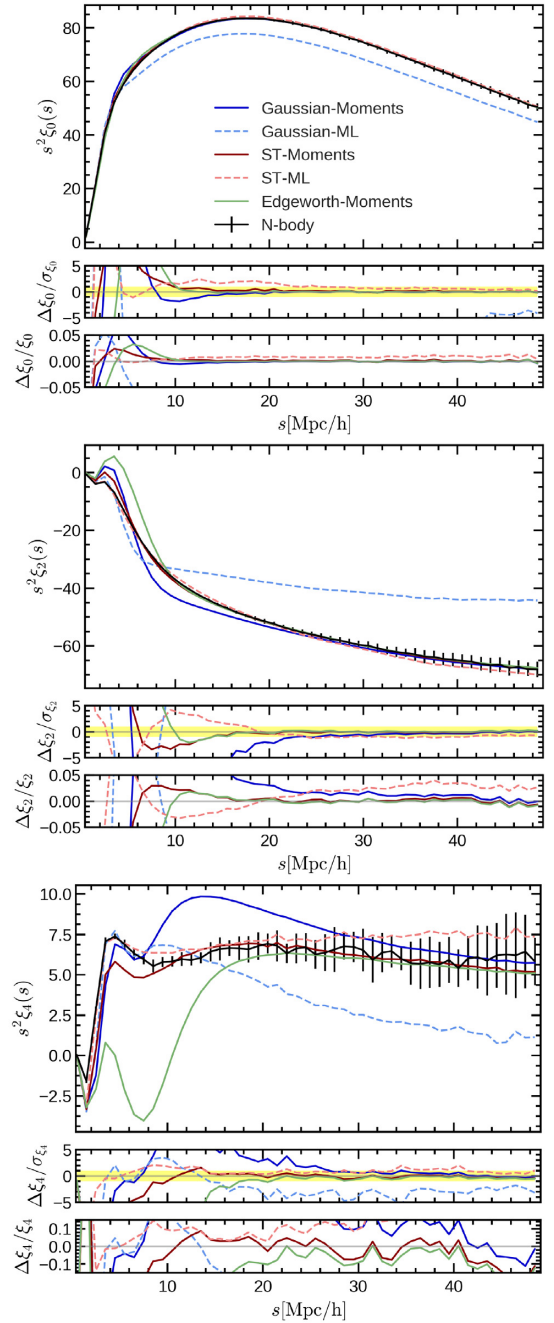


Figure 6. Comparison of the accuracy of the different models for reproducing the multipoles of the redshift space correlation function. In the upper sub-panels the multipole directly measured from the simulation is shown together with the model predictions. In the lower sub-panels the deviation between the model and the simulation in units of the variance calculated across the different independent simulations are shown. The yellow bands show the 1σ deviation.

Finally, although the measurement of the hexadecapole is itself very noisy, the ST model is within one standard deviation for scales larger than around $10 h^{-1}\text{Mpc}$, whilst the Gaussian model on those scales is already more than 5σ away from the measurement from the simulations.

To sum up, we have found that the use of the method of moments is critical to accurately reproduce the clustering on quasi-linear scales. The accuracy of the GSM we obtain is consistent with

previous findings (Reid & White 2011; Wang et al. 2014; Bianchi et al. 2016): the prediction is within the measurement errors from the simulations for scales larger than $30 h^{-1}\text{Mpc}$. However, the model prediction rapidly diverges from the simulation results on smaller scales. On the contrary, the ST model is able to reproduce the redshift space clustering very accurately on scales down to $10 h^{-1}\text{Mpc}$, by introducing a pairwise velocity distribution that incorporates the skewness and kurtosis of the pairwise velocity PDF.

On the other hand, we need to understand why the Gaussian model reproduces the clustering on scales above $30 h^{-1}\text{Mpc}$ more accurately than the ST distribution obtained through the ML method, even though the latter is a better description of the pairwise velocity distribution on those scales, as shown in Fig. 2. To this end, we will study the behaviour of the integrand of equation (6) in more detail in the next section.

4 THE IMPORTANCE OF THE MOMENTS FOR ACCURATE CLUSTERING PREDICTIONS

In this section, we show how the accuracy of the streaming model on quasi-linear scales is directly related to the lowest order moments of the pairwise velocity distribution. We start by studying how well the different models reproduce the streaming model integrand.

4.1 Lessons from the streaming model integrand

We show the integrand of equation (6), for a few pair separation vectors, in Fig. 7. In broad terms the integrand is the outcome of a competition between the probability of finding a pair of haloes at a given separation, i.e. the two-point correlation function, and the probability that the pair has the necessary relative velocity to move from real space position \mathbf{r} to redshift space position \mathbf{s} . Whilst the first quantity is evaluated as $\xi\left(\sqrt{s_{\perp}^2 + r_{\parallel}^2}\right)$ for fixed s_{\perp} , and therefore peaks at $r_{\parallel} \sim 0$, the latter is evaluated as $\mathcal{P}(v_{\parallel} = s_{\parallel} - r_{\parallel})$, and peaks around its mean, close to $v_{\parallel} \approx 0$ ($r_{\parallel} \approx s_{\parallel}$) for large pair separations. The effect of this competition can be seen in Fig. 7. For large pair separations, e.g. as shown by the middle and bottom panels, the real-space correlation function is small, $\xi\left(\sqrt{s_{\perp}^2 + r_{\parallel}^2}\right) \ll 1$, so that the integrand is dominated by $\mathcal{P}(v_{\parallel} = s_{\parallel} - r_{\parallel})$ and has a peak at $r_{\parallel} \approx s_{\parallel}$. On the other hand, for small pair separations (the top panel), ξ is no longer negligible and the integrand acquires a second, albeit smaller peak around $r_{\parallel} \approx 0$.

As for the different streaming models, the Gaussian one obtained through the method of moments systematically shifts the main peak of the integrand from its true position, and makes it wider. Although this seems to be a poorer estimate of the integrand than the Gaussian model obtained through maximizing the likelihood, which is consistent with what Fig. 2 suggests, it predicts the clustering multipoles with a precision that is one order of magnitude higher after integrating, as can be seen on the resulting redshift space correlation function annotated on Fig. 7. This same effect is present on all scales larger than $s \approx 30 h^{-1}\text{Mpc}$.

More interestingly, both the ST moments and the ST ML methods give visually much better predictions for the integrand than the Gaussian moments method, which is a consequence of the pairwise velocity distribution being non-Gaussian for all pair separations. However, as shown in the previous section, after integration we find that the Gaussian model yields a comparable accuracy to the non-Gaussian ST model on scales larger than $30 h^{-1}\text{Mpc}$ for the

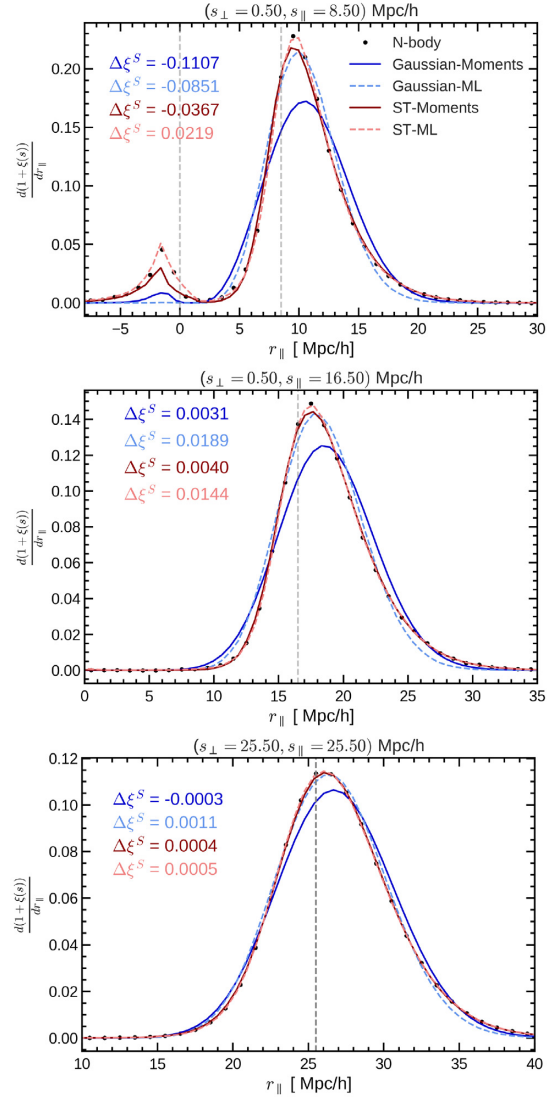


Figure 7. Integrand of equation (6) shown for different redshift space pair separations. At small pair separations and for small μ (top panel) we find two peaks situated at $r_{\parallel} = 0$ and $r_{\parallel} = s_{\parallel}$, marked by the grey vertical dashed lines. For larger μ (middle and bottom panels), the second peak dominates since the correlation function decays rapidly at large separations. The result of the integral for the different models minus the integral obtained using the pairwise velocity distribution measured from the simulations, $\Delta\xi^S(s)$, is plotted with different models shown by the different colours and line styles as shown by the legend in the top panel.

monopole and quadrupole. This coincidental behaviour has been noted previously by Kuruvilla & Porciani (2018). Taking the middle panel of Fig. 7 as an example, the ‘errors’ of the integration in equation (6), $\Delta\xi^S$, defined as the difference between the integration of the model curve and the integration using the simulation results (black dots), for the four streaming models considered here, are shown in the figure labels. We note that the Gaussian moments method gives a slightly smaller error than the ST moments at the particular pair separation shown. As the former underestimates the integrand for $14 \lesssim r_{\parallel}/(h^{-1}\text{Mpc}) \lesssim 20$ and $r_{\parallel} \gtrsim 26 h^{-1}\text{Mpc}$, and overestimates it in other regimes, this seems to suggest that a precise cancellation of the errors from different r_{\parallel} intervals takes place, which makes the final integration result accurate. However, this cancellation of errors happens for all larger pair separations

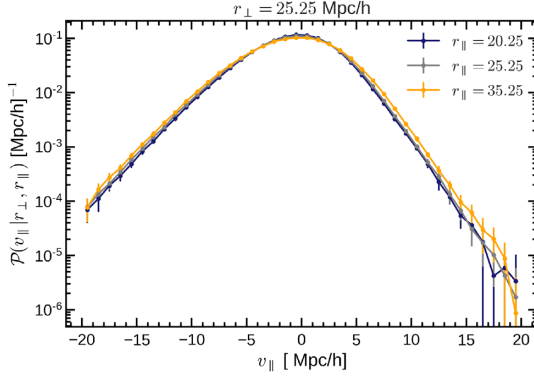


Figure 8. The dependence of the pairwise velocity distribution on r_{\parallel} , for fixed $r_{\perp} = 25.25 h^{-1} \text{Mpc}$ and for three values of r_{\parallel} over a range of $15 h^{-1} \text{Mpc}$ (20.25, 25.25, and $35.25 h^{-1} \text{Mpc}$). The distribution has only a weak dependence on r_{\parallel} , which is why the Taylor expansion described in the text works.

$s > 30 h^{-1} \text{Mpc}$. In the next subsection, we will show that this is a consequence of the integration being sensitive only to the moments of the pairwise velocity distribution. In particular, for large pair separations it is the two lowest order moments which dominate the outcome of the integral equation (6), while higher order moments only become important on scales smaller than $30 h^{-1} \text{Mpc}$ (Fig. 6).

4.2 The importance of the moments on quasi-linear scales

The integration in the streaming model, equation (6), is different from taking the expectation value of $1 + \xi^R(r)$ since the pairwise velocity distribution $\mathcal{P}(v_{\parallel}|\mathbf{r})$ is different for different r_{\parallel} values, rather than a fixed probability distribution function $\mathcal{P}(v_{\parallel})$. However, $\mathcal{P}(v_{\parallel}|\mathbf{r})$ is a slowly varying function of pair separation r_{\parallel} , for $r_{\parallel} \gtrsim 15 h^{-1} \text{Mpc}$, as can be seen in Fig. 8. The outcome of the streaming model integral for the values ($s_{\perp} = r_{\perp} = 25.5$, $s_{\parallel} = 25.5$), shown in the bottom panel of Fig. 7, is dominated by contributions from the pairwise velocity distribution in the range $20 h^{-1} \text{Mpc} < r_{\parallel} < 35 h^{-1} \text{Mpc}$, which is the range of values shown in Fig. 8. The same feature is found at other separations larger than about $10 h^{-1} \text{Mpc}$.

Therefore, we can Taylor expand the integrand around its peak at $r_{\parallel} = s_{\parallel}$ as follows:

$$(1 + \xi^R(r)) \mathcal{P}(v_{\parallel}|\mathbf{r}) \approx (1 + \xi^R(s)) \mathcal{P}(v_{\parallel}|s) + \sum_n \frac{1}{n!} (r_{\parallel} - s_{\parallel})^n \frac{d^n}{ds_{\parallel}^n} ((1 + \xi^R(s)) \mathcal{P}(v_{\parallel}|s)). \quad (21)$$

This expansion was already used by Fisher (1995), Scoccimarro (2004), and Bianchi et al. (2015) to obtain the Kaiser limit of the streaming model. Here, we will show that this is still accurate on quasi-linear scales.

Inserting equation (21) into equation (6), we find that the derivatives with respect to s_{\parallel} can be taken out of the integral over r_{\parallel} , together with the real space correlation function, and therefore after integration we are left with the derivatives of the moments through the dependency of v_{\parallel} on r_{\parallel} . For the lowest order term on the right-hand side of equation (21), we find after a change of variables $v_{\parallel} = s_{\parallel} - r_{\parallel}$,

$$\int_{-\infty}^{\infty} dv_{\parallel} (1 + \xi^R(s)) \mathcal{P}(v_{\parallel}|s) = (1 + \xi^R(s)), \quad (22)$$

whilst for the higher order terms

$$\begin{aligned} & \int_{-\infty}^{\infty} dv_{\parallel} (-1)^n \sum_n \frac{1}{n!} v_{\parallel}^n \frac{d^n}{ds_{\parallel}^n} ((1 + \xi^R(s)) \mathcal{P}(v_{\parallel}|s)) \\ &= \sum_n \frac{(-1)^n}{n!} \frac{d^n}{ds_{\parallel}^n} \left((1 + \xi^R(s)) \int_{-\infty}^{\infty} dv_{\parallel} v_{\parallel}^n \mathcal{P}(v_{\parallel}|s) \right) \\ &= \sum_n \frac{(-1)^n}{n!} \frac{d^n}{ds_{\parallel}^n} ((1 + \xi^R(s)) m_n(s)), \end{aligned} \quad (23)$$

where m_n denotes the n th order moment about the origin of the pairwise velocity distribution, which is related to the central moments through

$$m_n = \sum_{k=0}^n \binom{n}{k} c_k m_1^{n-k}. \quad (24)$$

As a result, an approximation to the streaming model is given by

$$\xi^S(s_{\perp}, s_{\parallel}) \approx \xi^R(s) + \sum_n \frac{(-1)^n}{n!} \frac{d^n}{ds_{\parallel}^n} ((1 + \xi^R(s)) m_n(s)), \quad (25)$$

where the integral of equation (6) has now been replaced by derivatives of the pairwise velocity moments evaluated at the redshift space position, \mathbf{s} . Consequently, for large pair separations, where the above approximation works well, the exact shape of the pairwise velocity distribution does not affect the clustering, and it is only the moments of the distribution that influence the redshift space correlation function. This explains why the Gaussian moments model works so well in Fig. 6 while the Gaussian ML model, which describes the integrand better, fails to reproduce the multipoles.

We can use equation (25) to obtain analytical predictions for the redshift space clustering based on the moments. Up to first-order terms the resulting expression is simply

$$\begin{aligned} \xi^{(1)}(s, \mu) &\approx \xi^R(s) - \frac{d\xi^R(s)}{ds} m_{10}(s) \mu^2 \\ &\quad - (1 + \xi^R(s)) \left(\frac{m_{10}(s)}{s} (1 - \mu^2) + \frac{dm_{10}(s)}{ds} \mu^2 \right), \end{aligned} \quad (26)$$

where $m_{10} = m_{10}(s)$, defined by equation (16), denotes the radial mean infall. For the multipoles we have

$$\xi_0^{(1)}(s) \approx \xi^R - \frac{1}{3} \frac{d\xi^R}{ds} m_{10} - \frac{1}{3} (1 + \xi^R) \left(2 \frac{m_{10}}{s} + \frac{dm_{10}}{ds} \right), \quad (27)$$

$$\xi_2^{(1)}(s) \approx -\frac{2}{3} \frac{d\xi^R}{ds} m_{10} + \frac{2}{3} (1 + \xi^R) \left(\frac{m_{10}}{s} - \frac{dm_{10}}{ds} \right), \quad (28)$$

$$\xi_4^{(1)}(s) \approx 0. \quad (29)$$

For second-order terms, we find

$$\begin{aligned} \xi^{(2)}(s, \mu) &\approx \frac{1}{2} \left(\frac{d^2 \xi(s)}{ds_{\parallel}^2} m_2(s_{\parallel}, s_{\perp}) + (1 + \xi(s)) \frac{d^2 m_2(s_{\parallel}, s_{\perp})}{ds_{\parallel}^2} \right) \\ &\quad + \frac{d\xi(s)}{ds_{\parallel}} \frac{dm_2(s_{\parallel}, s_{\perp})}{ds_{\parallel}}. \end{aligned} \quad (30)$$

Note that the analytical result for the second-order multipoles includes many more terms than its first-order equivalent. Therefore, instead of calculating the resulting multipoles analytically, we take numerical derivatives of moments higher than one in the remainder of this work.

In what follows we address the question of how accurate the expansion equation (25) is on small scales, and how the different moments affect the clustering multipoles. The expansion turns out

to give accurate predictions for the multipoles even on scales of about $10 h^{-1}\text{Mpc}$, with the advantage of replacing the integral in equation (6), which sums up the contributions of the pairwise velocity distributions on different scales, with a derivative at the scale s under consideration. Therefore, it converts the non-local relationships between the redshift and real space correlation functions with the pairwise velocity PDF, into a local relation between the redshift space correlation and the derivatives of the pairwise velocity moments and real space correlation function.

In the next subsection, we will test the accuracy of the expansion by comparing it to a full streaming model in which we assume the pairwise velocity distribution is either Gaussian or ST.

4.3 The range of validity of the streaming model expansion

Assessing the exactitude of the Taylor expansion on different scales is not straightforward, since including higher order terms involves higher order derivatives of the velocity moments and the real space correlation function.

As one can see from Fig. 4, on small scales the velocity moments as functions of pair separation measured from the simulation are not smooth and their high-order derivatives can be noisy, which will affect the accuracy of the analytical predictions. Given that our main objective in this subsection is to test the validity of the expansion method, to eliminate the impact of such noise, we fit the ingredients of the Taylor expansion of the streaming model. The fits to the moments are shown as dotted lines in Fig. 4. For the real space correlation function, we fit with a simple power law. We then take the fitted curves as the ‘truth’, and compare the predictions of the Taylor expansion with two different streaming models in which we can convert the fitted moments into pairwise velocity distributions. These include a GSM, to demonstrate why the GSM gives accurate predictions on quasi-linear scales where the pairwise velocity distribution is highly non-Gaussian, and an ST model, to show the effect of skewness and kurtosis in improving the accuracy of the expansion. The Gaussian distribution has the correct two lowest central moments, whilst the ST matches the four lowest moments.

We shall not compare the expansion to the simulation measurements directly, since the analytical fitting formulae to the measured moments already induce at least per cent-level modifications to the multipoles on small scales. However, this exercise is realistic enough (both the real space correlation function and the moments have been fitted to the N -body simulation results) to demonstrate up to which scale the Taylor expansion method can be used.

Compared with the full streaming model, the Taylor expansion makes minimal assumptions regarding the pairwise velocity distribution, since it only uses the moments, and removes the integration over all pair separations in equation (6). For the Gaussian case, the results of expanding equation (25) up to $n = 4$ are shown as coloured lines in Fig. 9. Note that although for a Gaussian distribution the odd central moments vanish, the odd moments about the origin get contributions from lower order even central moments, as expressed in equation (24). Therefore, even orders of the expansion do contribute. Similarly, while odd central moments higher than the second order vanish for a Gaussian distribution, m_n is non-zero for $n > 2$. The full streaming model predictions using the integral in equation (6) and a Gaussian pairwise velocity distribution are shown as black solid lines.

In Fig. 9, we see how including only terms up to $n = 2$ we can reproduce the monopole to within 1 per cent down to $10 h^{-1}\text{Mpc}$. To achieve a comparable accuracy for the quadrupole, however, we need to add higher order moments up to $n = 4$. For the hexadecapole,

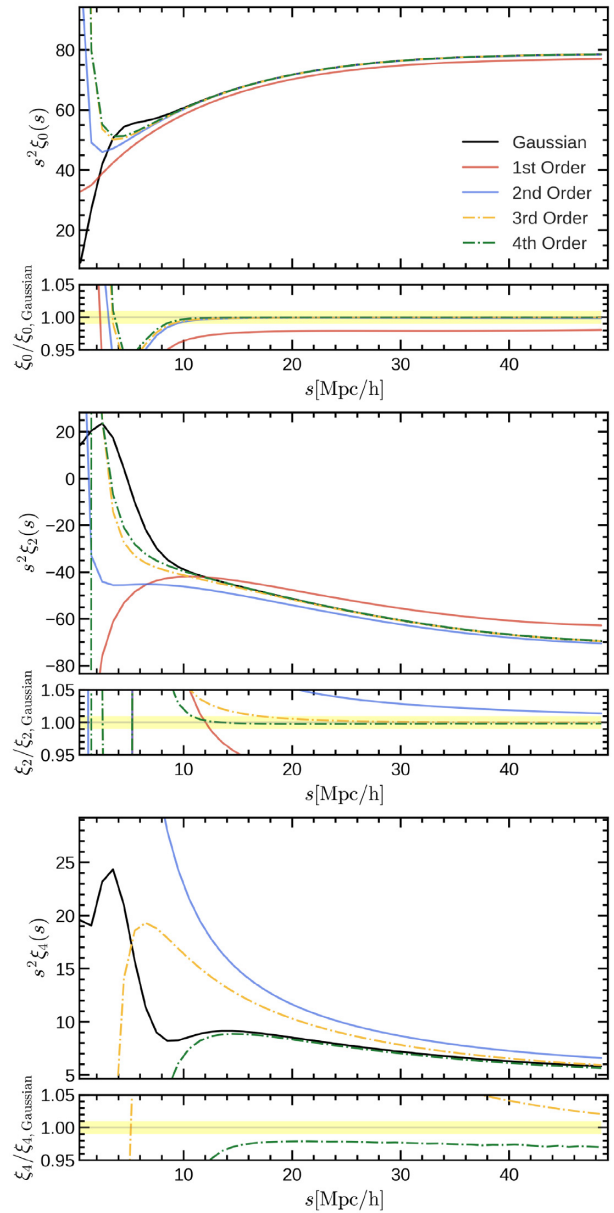


Figure 9. Accuracy of the Taylor expansion up to fourth order, assuming the pairwise velocity distribution is Gaussian, compared to the full streaming model under the same assumptions. Note all the real space ingredients to the streaming model, the real space correlation function and the two lowest order pairwise velocity moments, are analytical functions fitted to the simulation measurements. The yellow shaded region shows one per cent level agreement between the Taylor expansion and the full streaming model. The monopole achieves an accuracy better than the one per cent on scales above $10 h^{-1}\text{Mpc}$ when the expansion is truncated at second order. For the quadrupole to achieve a similar accuracy, we need to retain up to fourth-order terms.

we expect the Taylor expansion to be less accurate, because it is more strongly affected by the finger-of-God effect, which originates from the very small and non-linear scales on which the Taylor expansion breaks down. This is confirmed by the fact that in the lower panel of Fig. 9 there are larger differences between the coloured and black lines. Nevertheless, we find that for the Gaussian model, the hexadecapole predicted by the expansion up to fourth order is accurate to within 3 per cent down to $15 h^{-1}\text{Mpc}$.

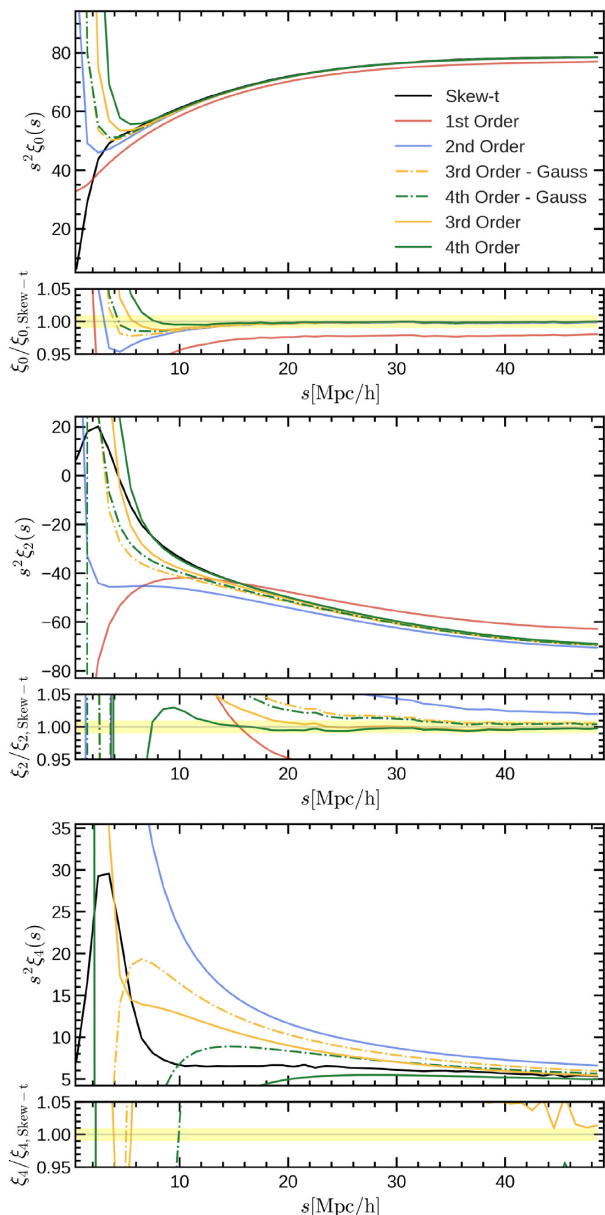


Figure 10. Same as in Fig. 9, but assuming the pairwise velocity distribution follows an ST distribution. Compared with a Gaussian distribution, which has the correct first- and second-order moments, the ST also matches the skewness and kurtosis. For comparison, we show the Taylor expansion found assuming the distribution is Gaussian and the fully non-Gaussian result, in which we include skewness and kurtosis. The effect of the skewness can therefore be seen as the difference between the orange dashed and orange solid lines, whilst the effect of the kurtosis is given by the difference between the green dashed and green solid lines. We find that the effects of the skewness and kurtosis are particularly important for the quadrupole on small scales.

Finally, since the ST distribution reproduces the measured line-of-sight velocity distribution with a higher accuracy than a Gaussian distribution (Fig. 2), we also demonstrate the effect of higher order moments on the Taylor expansion using an ST model for \mathcal{P} . In Fig. 10, we show both the third- and fourth-order moment expansion assuming a Gaussian distribution, with zero skewness and fixed kurtosis shown as dash-dotted lines, and the fully non-Gaussian moments. Although for the monopole, non-Gaussianity does not play an important role, adding the skewness extends the 1 per cent

agreement in the quadrupole from scales of around $30 h^{-1} \text{Mpc}$ down to $20 h^{-1} \text{Mpc}$. The effect of the fourth-order moment, kurtosis, is important to extend close agreement even further to about $10 h^{-1} \text{Mpc}$. These results are consistent with the findings in Fig. 6, where we find that the ST model improves the agreement of the quadrupole in the range $10\text{--}30 h^{-1} \text{Mpc}$.

Note that for the hexadecapole (shown in the bottom panel of Fig. 10) the Taylor expansion method introduces a substantial fractional error of > 5 per cent on all scales, even if the fourth-order corrections are included. This is not surprising because now the assumed true model – in which the pairwise velocity satisfies an ST distribution – is more complicated, and because the absolute value of the hexadecapole is much closer to zero which tends to magnify the relative error. Nevertheless, we still observe that including higher order terms brings the expansion prediction closer to the correct answer. In Fig. 6, the multipoles have been numerically calculated using the full streaming model of equation (6), rather than the Taylor expansion, under the assumption of the pairwise velocity PDF being either Gaussian or ST.

5 SENSITIVITY TO THE REAL SPACE STREAMING MODEL INGREDIENTS

In this section, we study the effects of varying the various ingredients – the real space quantities – needed to predict the redshift space clustering through the ST streaming model. This will indicate to us what precision is required for each ingredient in order to make the final prediction of the multipoles accurate.

In Fig. 11, we show the effects on the multipoles of varying the real space correlation function (the first row), the mean pairwise velocity (the second row) and its variance (both in the radial and transverse components; the last two rows). We have studied the impact of two types of variations: a constant change by ± 5 per cent, or a fractional change that increases towards smaller pair distances as $1/r$, to emulate the fact that perturbation theory predictions worsen towards small scales. The latter gradual change is tuned to vary the given function by ± 5 per cent on scales of $5 h^{-1} \text{Mpc}$ and by ± 1 per cent on the scale of $30 h^{-1} \text{Mpc}$. In this way, we can compare the effect of a varying slope due to uncertainties in our predictions, which we know is important since derivatives of the moments appear in the Taylor expansion equation (25). The fractional changes to the predicted redshift space monopole, quadrupole and hexadecapole (from left to right) are respectively shown as orange and blue shaded regions for the constant and scale-dependent changes.

Varying the real space correlation function by 5 per cent produces approximately the same fractional change in the monopole. Since the zeroth order contribution to the quadrupole and hexadecapole is zero, both the constant and the scale-dependent variations in the real space correlation function produce a sub-per cent effect on the quadrupole and hexadecapole on scales larger than $20 h^{-1} \text{Mpc}$.

The mean pairwise velocity has a stronger effect on both the quadrupole and hexadecapole. The importance of getting the slope of the mean right, found in the Taylor expansion, can also be seen in the blue contours. A change of 1 per cent above $30 h^{-1} \text{Mpc}$ produces a slightly larger effect on the quadrupole. The monopole, in contrast, is much less sensitive to variations of the mean pairwise velocity, and the effect is at the sub-per cent level at $s \gtrsim 10 h^{-1} \text{Mpc}$ for both variation scenarios.

On the other hand, the hexadecapole is most sensitive to the radial and transverse standard deviations. Changes of 5 per cent can produce a change that is twice as large in the hexadecapole.

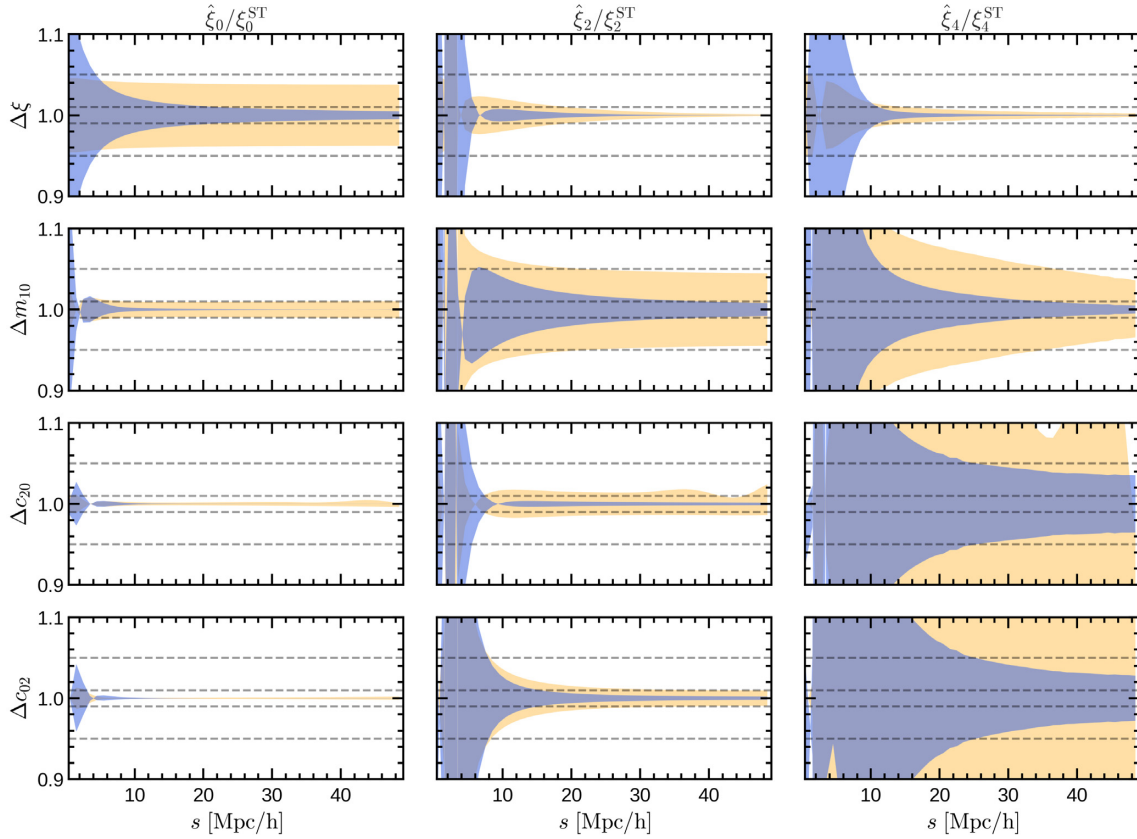


Figure 11. The fractional variation in the monopole, quadrupole, and hexadecapole after modifying the real space ingredients of the streaming model. In each row, we show the effect of varying: the real space correlation function, the mean pairwise velocity, the radial variance of the pairwise velocity and the transverse variance of the pairwise velocity. Orange contours show the effect of varying each of these quantities by ± 5 per cent, while the blue contours vary them by a percentage that depends on scale and increases with $1/r$. On small scales, where perturbation theory predictions degrade, we vary each of the ingredients by a larger percentage. The variation is tuned to produce a 5 per cent change on scales of $5 h^{-1} \text{Mpc}$ and a 1 per cent one on scales of $30 h^{-1} \text{Mpc}$. Gray dashed lines determine 5 and 1 per cent deviations from the true model.

Regarding the third-order moments, we found in the last section (Fig. 10) that the skewness has at most a per cent-level effect on the monopole and quadrupole on scales below $30 h^{-1} \text{Mpc}$. Since its effect is very small, we do not show the equivalent in Fig. 11. We find that varying the third-order radial and transverse moments by 50 per cent introduces modifications smaller than 5 per cent on the quadrupole on small scales.

Finally, the effect of fourth order terms is important on scales below $20 h^{-1} \text{Mpc}$. However, we have already shown in the previous section that setting the fourth-order moments to zero, by assuming Gaussianity, also gives only a few percentage level corrections to the quadrupole on small scales (see difference between orange dashed line and solid green line in Fig. 10).

Therefore, even on small scales, we need to predict most accurately the lower order moments: the mean and the standard deviation, particularly the latter, if we want to utilize information contained in the hexadecapole. We can afford to have a larger margin of error on the predictions of the higher order moments, and still extend the validity up to scales of around $10 h^{-1} \text{Mpc}$.

6 CONCLUSIONS AND DISCUSSION

The new generation of surveys (Amendola et al. 2013; Takada et al. 2014; de Jong et al. 2018; Levi et al. 2019) is going to measure redshift space clustering of galaxies with unprecedented precision.

To translate the high accuracy of these measurements into tighter constraints on the cosmological parameters or on possible deviations from general relativity, we need to improve our theoretical models of RSD. Within the streaming model of RSD, we need to: (i) improve the mapping from real to redshift space, i.e. by developing the modelling of the pairwise velocity distribution including its higher order moments, (ii) increase the accuracy of the predictions of the ingredients of the streaming model – the real space correlation function and the pairwise velocity moments – for given cosmological parameters. Here, we have focused on the first of these aspects, but we have also briefly analysed the effects of the second.

In N -body simulations, where the fully non-linear evolution of collisionless particles is solved, we observe that the distribution of the pairwise velocities of dark matter haloes is skewed towards negative velocities, and has broader tails than a Gaussian. Therefore, models that use Gaussian distributions do not give an accurate description of the pairwise velocities. We have introduced an extension to the GSM by using the ST probability distribution for the pairwise velocity. The parameters of this distribution can be tuned to match the four lowest order velocity moments measured from simulations. The ST model describes the simulation measurement of the pairwise velocity distribution significantly better than a simple Gaussian.

We compare two different methods to find the best-fitting parameters of the pairwise velocity distribution: maximum likelihood estimation and the method of moments. Although the results of

both approaches seem to describe the measured velocity distribution equally well on large scales, they give very different results for the redshift space clustering once inserted into the streaming model. Using the method of moments is crucial for describing all multipoles, including the small scales. Even though the Gaussian distribution gives a very poor fit to the measured pairwise velocities distribution, it can reproduce the true multipoles on quasi-linear scales within the small statistical errors of our simulations when we tune it to have the two lowest order moments extracted from the simulations. On the other hand, the best-fitting Gaussian found by maximizing the likelihood gives results that are more than five standard deviations away from the simulation measurement.

The ST model, also using the method of moments, gives predictions for the redshift space multipoles (monopole, quadrupole, and hexadecapole) that are within the small statistical sampling variance errors (driven by the simulation volume) down to about $10 h^{-1}\text{Mpc}$. On such small scales, the GSM gives predictions that are more than five standard deviations away from the mean measurement from simulations. Therefore, the ST model extends the validity of the streaming model from 30 to $10 h^{-1}\text{Mpc}$, and gives a more accurate description of the hexadecapole, which has so far not been used in analyses that rely on the GSM (e.g. Satpathy et al. 2017; Zarrouk et al. 2018), due to its poor accuracy.

We have used a Taylor expansion of the integrand to show why the GSM can reproduce the clustering on quasi-linear scales within the error bars of the simulation measurement, despite giving a poor description of the pairwise velocity distribution. At $s \gtrsim 30 h^{-1}\text{Mpc}$, only the first- and second-order moments, the mean and the standard deviation, of the pairwise velocity distribution, are crucial for determining the monopole and quadrupole of the two-point correlation function in redshift space.

We have also shown that the Taylor expansion can describe the non-Gaussian ST streaming model down to smaller scales, of about $10 h^{-1}\text{Mpc}$, when expanded up to fourth order. The main advantage of the Taylor expansion is that it replaces the integral of the pairwise velocities over all scales by a derivative of the moments at the scale under consideration. It therefore makes no assumptions about the details of the underlying velocity distribution, and can give analytical predictions for the monopole and quadrupole. However, it cannot reproduce the hexadecapole as accurately as the full ST streaming model integral, equation (6), or is it as accurate on smaller scales, $s \lesssim 15 h^{-1}\text{Mpc}$.

The Taylor expansion could be particularly useful to measure the velocity moments from the observed redshift space multipoles, as was already proposed by Bianchi et al. (2015), along the line of previous measurements of the pairwise velocity dispersion (Li et al. 2006; Loveday et al. 2018). The main difficulty to measure the pairwise distribution from observations lies in the pair distance dependence of the moments, imprinted by gravity. We would need to develop analytical formulae to summarize the pair distance dependence in a small set of parameters that are valid independently of the underlying model of gravity. These parameters could then be inferred from observations of redshift space clustering, by running a Markov chain Monte Carlo. The direct measurement of the moments could be a complementary test of gravity to the growth rate, and it would utilize more information of the full scale dependence of different gravity models.

Finally, we qualitatively analysed the effects of inaccurate knowledge of the real space correlation function or of the velocity moments on the predictions of the redshift multipoles. As expected, the monopole is mainly determined by the real space correlation function. We have shown that perturbation theory based CLEFT method

(Vlah et al. 2016) produces per cent-level-accuracy predictions of the real space correlation function. However, to obtain per cent-level accurate predictions for both the monopole and quadrupole, we also need similar accuracy for the mean pairwise velocity and its slope. Fitting the CLEFT predictions, with five free parameters, we were only able to obtain predictions accurate at the per cent level for the mean on scales above $35 h^{-1}\text{Mpc}$. On the other hand, the hexadecapole is very sensitive to the variance of pairwise velocities, for which CLEFT is only accurate to one per cent above scales of about $45 h^{-1}\text{Mpc}$. Therefore, future efforts to utilize the information content in the hexadecapole will have to obtain more accurate theoretical prescriptions for the variance. Per cent level errors on the prediction of the variance become even larger errors on the hexadecapole. On scales smaller than $30 h^{-1}\text{Mpc}$, we also need predictions for the skewness and kurtosis of pairwise velocities. However, these do not need to be as accurate: per cent errors on the skewness and kurtosis have negligible impact on the multipoles.

To summarize, we have developed a streaming model based on the ST distribution of pairwise velocities, that accurately describes redshift space clustering on scales larger than $10\text{--}15 h^{-1}\text{Mpc}$, given the first four moments of the pairwise velocity distribution are known. In order to improve constraints on the growth rate by using the ST model, we need to improve the theoretical predictions of the pairwise velocity moments and its dependency on the cosmological parameters. In this work, we have focused our analysis on massive dark matter haloes at redshift zero, and we leave the study of galaxies at a range of different redshifts to future work.

ACKNOWLEDGEMENTS

We would like to thank Davide Bianchi, Ravi Sheth, Emanuele Castorina, Joseph Kuruvilla, and Cristiano Porciani for useful discussions. CCL is supported by a PhD Studentship from the Durham Centre for Doctoral Training in Data Intensive Science, funded by the UK Science and Technology Facilities Council (STFC, ST/P006744/1) and Durham University. BL, PZ, and CMB acknowledge support from the STFC through ST/P000541/1. BL and AE acknowledge the support of the European Commission through the European Research Council (ERC-StG-PUNCA-716532). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, and ST/R002371/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. This work was supported in part by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, JSPS KAKENHI Grant Numbers JP15H03654, JP15H05887, JP15H05893, JP15H05896, JP15K21733, JP17K14273, and JP19H00677, and Japan Science and Technology Agency CREST JPMHCR1414. TN and MT acknowledge the supported by the Munich Institute for Astro- and Particle Physics (MIAPP) of the DFG cluster of excellence ORIGINS (<http://www.munich-iapp.de>). Numerical simulations presented in this paper were carried out on Cray XC30 and XC50 at Centre for Computational Astrophysics, National Astronomical Observatory of Japan.

DATA AVAILABILITY

The data shown in this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Abidi M. M., Baldauf T., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 029
- Amendola L. et al., 2013, *Living Rev. Relativ.*, 16, 6
- Azzalini A., Capitanio A., 2003, *J. R. Stat. Soc. B*, 65, 2
- Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, *Phys. Rev. D*, 86, 083540
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, *ApJ*, 762, 109
- Bianchi D., Chiesa M., Guzzo L., 2015, *MNRAS*, 446, 75
- Bianchi D., Percival W. J., Bel J., 2016, *MNRAS*, 463, 3783
- Bose B., Koyama K., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 029
- Carlson J., Reid B., White M., 2013, *MNRAS*, 429, 1674
- Carrasco J. J. M., Hertzberg M. P., Senatore L., 2012, *J. High Energy Phys.*, 2012, 82
- Carroll S. M., 2001, *Living Rev. Relativ.*, 4, 1
- Chan K. C., Scoccimarro R., Sheth R. K., 2012, *Phys. Rev. D*, 85, 083509
- Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, *Phys. Rep.*, 513, 1
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- de Jong R. S. et al., 2018, in Evans C. J., Simard L., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII. SPIE, Bellingham
- Fisher K. B., 1995, *ApJ*, 448, 494
- Gagrani P., Samushia L., 2017, *MNRAS*, 467, 928
- Gorski K., 1988, *ApJ*, 332, L7
- Guzzo L. et al., 2008, *Nature*, 451, 541
- Hamilton A. J. S., 1998, in Hamilton D., ed., Astrophysics and Space Science Library, Vol. 231, The Evolving Universe. Springer-Verlag, Berlin, p. 185
- Hernández-Aguayo C., Hou J., Li B., Baugh C. M., Sánchez A. G., 2019, *MNRAS*, 485, 2194
- Huterer D. et al., 2015, *Astropart. Phys.*, 63, 23
- Jennings E., Baugh C. M., Pascoli S., 2011, *ApJ*, 727, L9
- Joyce A., Lombriser L., Schmidt F., 2016, *Ann. Rev. Nucl. Part. Sci.*, 66, 95
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kuruvilla J., Porciani C., 2018, *MNRAS*, 479, 2256
- Lazeyras T., Schmidt F., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 008
- Lazeyras T., Wagner C., Baldauf T., Schmidt F., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 018
- Levi M. E. et al., 2019, preprint (arXiv:1907.10688)
- Li C., Jing Y. P., Kauffmann G., Börner G., White S. D. M., Cheng F. Z., 2006, *MNRAS*, 368, 37
- Loveday J. et al., 2018, *MNRAS*, 474, 3435
- Matsubara T., 2008, *Phys. Rev. D*, 78, 083519
- Nishimichi T. et al., 2019, *ApJ*, 884, 29
- Perlmutter S. et al., 1997, *ApJ*, 483, 565
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Planck Collaboration VI, 2018, *A&A*, in press, doi:10.1051/0004-6361/201833910
- Reid B. A., White M., 2011, *MNRAS*, 417, 1913
- Riess A. G. et al., 1998, *AJ*, 116, 1009
- Riess A. G. et al., 2007, *ApJ*, 659, 98
- Samushia L., Percival W. J., Raccanelli A., 2012, *MNRAS*, 420, 2102
- Satpathy S. et al., 2017, *MNRAS*, 469, 1369
- Scoccimarro R., 2004, *Phys. Rev. D*, 70, 083007
- Sellentin E., Jaffe A. H., Heavens A. F., 2017, preprint (arXiv:1709.03452)
- Sheth R. K., 1996, *MNRAS*, 279, 1310
- Song Y.-S., Taruya A., Oka A., 2015, *J. Cosmol. Astropart. Phys.*, 2015, 007
- Springel V., 2005, *MNRAS*, 364, 1105
- Takada M. et al., 2014, *PASJ*, 66, R1
- Taruya A., Nishimichi T., Saito S., 2010, *Phys. Rev. D*, 82, 063522
- Taruya A., Saito S., Nishimichi T., 2011, *Phys. Rev. D*, 83, 103527
- Tinker J. L., 2007, *MNRAS*, 374, 477
- Uhlemann C., Kopp M., Haugg T., 2015, *Phys. Rev. D*, 92, 063004
- Vlah Z., Castorina E., White M., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 007
- Wang L., Reid B., White M., 2014, *MNRAS*, 437, 588
- Zarrouk P. et al., 2018, *MNRAS*, 477, 1639
- Zu Y., Weinberg D. H., 2013, *MNRAS*, 431, 3319

APPENDIX A: METHOD OF MOMENTS FOR THE ST DISTRIBUTION

The four parameters of the ST distribution (v_c , w , α , ν) are determined by the first four order moments. To simplify the relation between moments and parameters, we introduce,

$$b_\nu = \left(\frac{\nu}{\pi}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma(\nu/2)}, \quad (\text{A1})$$

$$\delta = \frac{\alpha}{\sqrt{(1+\alpha^2)}}, \quad (\text{A2})$$

The moments are therefore given by

$$m_1 = v_c + w\delta b_\nu, \quad (\text{A3})$$

$$c_2 = w^2 \left(\frac{\nu}{\nu-2} - \delta^2 b_\nu^2 \right), \quad (\text{A4})$$

$$\gamma_1 = \frac{c_3}{c_2^{3/2}} = \delta b_\nu \left(\frac{\nu(3-\delta^2)}{\nu-3} - \frac{3\nu}{\nu-2} + 2\delta^2 b_\nu^2 \right) \left(\frac{\nu}{\nu-2} - \delta^2 b_\nu^2 \right)^{-3/2}, \quad (\text{A5})$$

$$\gamma_2 = \frac{c_4}{c_2^2} = \left(\frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4\delta^2 b_\nu^2 \nu(3-\delta^2)}{\nu-3} - \frac{6\delta^2 b_\nu^2 \nu}{\nu-2} - 3\delta^4 b_\nu^4 \right) \left(\frac{\nu}{\nu-2} - \delta^2 b_\nu^2 \right)^{-2}. \quad (\text{A6})$$

The parameters α and ν are obtained from the last two equations that determine the skewness and kurtosis of the distribution, these form a system of non-linearly coupled equations that we solve numerically. The remaining two parameters, v_c and w , can then directly be obtained from the equation for the mean and the variance.

APPENDIX B: ZOOM IN DISTRIBUTIONS

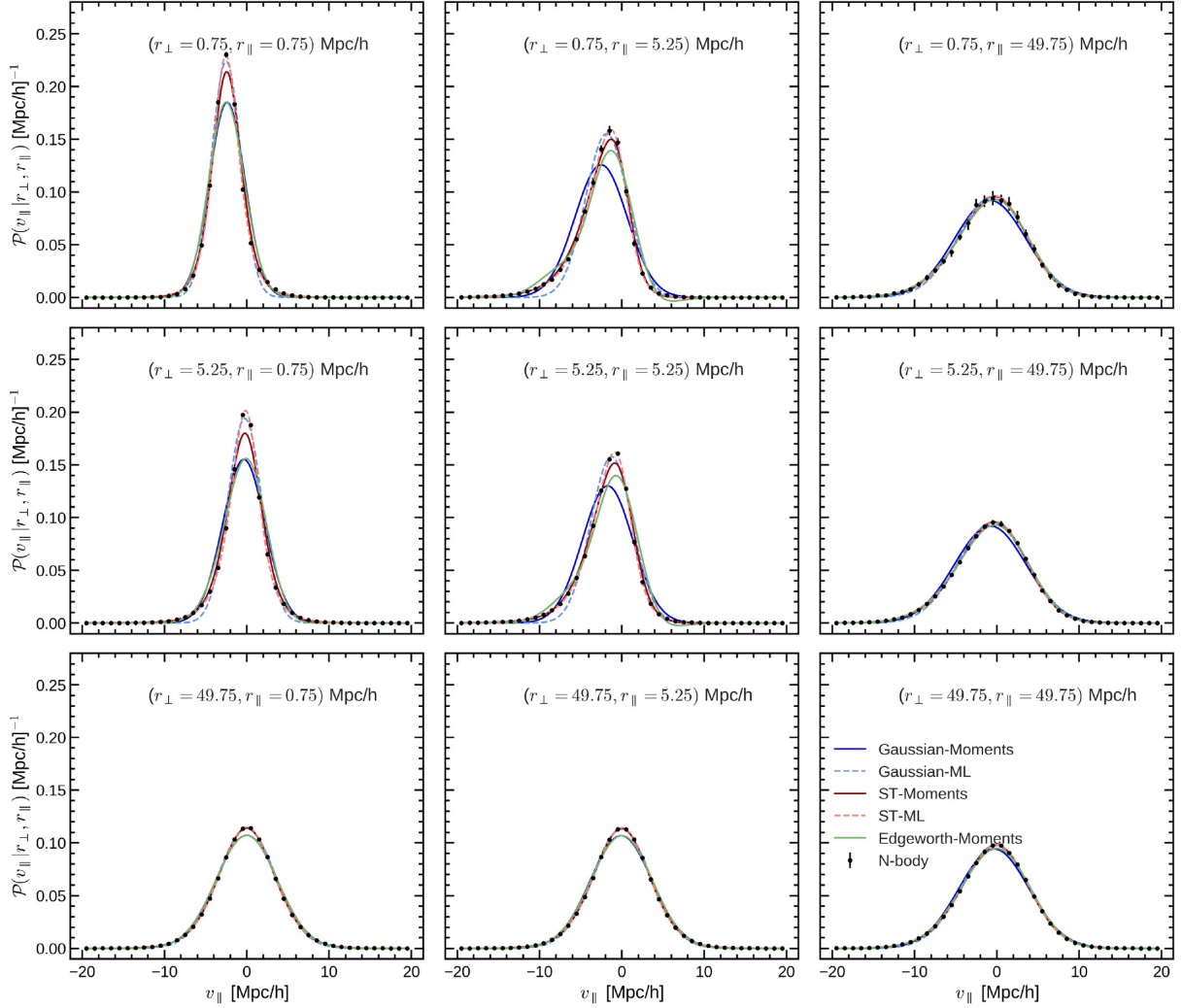


Figure B1. Linear scale representation of the pairwise velocity distribution to highlight the behaviour of the PDF close to its peak. The models shown are the same one as in Fig. 2. Note that the Edgeworth expansion predicts negative probabilities for certain pair separations such as $(r_{\perp} = 0.75, r_{\parallel} = 5.25) h^{-1} \text{Mpc}$ and $(r_{\perp} = 5.25, r_{\parallel} = 5.25) h^{-1} \text{Mpc}$, where the skewness is more pronounced. Moreover, an Edgeworth expansion behaves very differently from a Taylor expansion since it produces an asymptotic expansion, and therefore adding more terms does not guarantee convergence. See Sellentin et al. (2017) for an interesting discussion on the Edgeworth expansion and its applications to cosmology. In the application of the Edgeworth expansion to the pairwise velocity distribution, we see that it does not reproduce the N -body measurements as well as the ST distribution does with only one extra parameter.

APPENDIX C: PERTURBATION THEORY RESULTS IN DETAIL

In this Appendix, we show a detailed summary of the state-of-the-art CLPT and CLEFT perturbation theory predictions for the GSM ingredients. Note that we show the predictions for real space statistics, since we want to separately analyse the accuracy of perturbation theory predicting the ingredients of the streaming model, and the assumption of a Gaussian pairwise velocity distribution.

The free parameters are found by maximizing the combined Gaussian likelihood that the simulation measurements are most probable under the given theory

$$\log(\mathcal{L}) = \log(\mathcal{L}_\xi) + \log(\mathcal{L}_{m_{10}}) + \log(\mathcal{L}_{c_{20}}) + \log(\mathcal{L}_{c_{02}}), \quad (\text{C1})$$

where the individual likelihoods are given by

$$\log(\mathcal{L}_y) = -\frac{1}{2} \sum_i \frac{(y_{i,\text{measured}} - y_{i,\text{model}})^2}{\sigma_i^2}, \quad (\text{C2})$$

where y is the mean simulation measurement across the 15 independent simulations, and σ its standard deviation. Note that the covariance matrix is assumed to be diagonal, which means that the parameter uncertainties obtained from the fit will be underpredicted. While this assumption will also affect the values of the best-fitting parameters in detail, we do not expect this to have a qualitative impact on the relative agreement between the model predictions and data, which is our main objective here. We maximize the likelihood in the pair separation range $15 h^{-1}\text{Mpc} < r < 150 h^{-1}\text{Mpc}$ and the resulting mean parameter values are shown in Table C1. We find a value for the second-order Lagrangian bias b_2 that is in good agreement with previous measurements (Lazeyras et al. 2016), whereas the tidal bias is rather different from its local Lagrangian value ($b_s = 0$), which is in contrast with other analyses in the literature (Abidi & Baldauf 2018; Lazeyras & Schmidt 2018). We also note that the EFT parameters are the least constrained by our measurements, which is to be expected as they only have an impact on the small-scale regime.

In Fig. C1, we show a detailed comparison of the best-fitting model predictions for the two methods. The second counter-term introduced in CLEFT improves notably the prediction for the mean pairwise velocities on scales between 20 and 60 $h^{-1}\text{Mpc}$. Regarding the second-order moments, the predictions for m_{20} are similar for

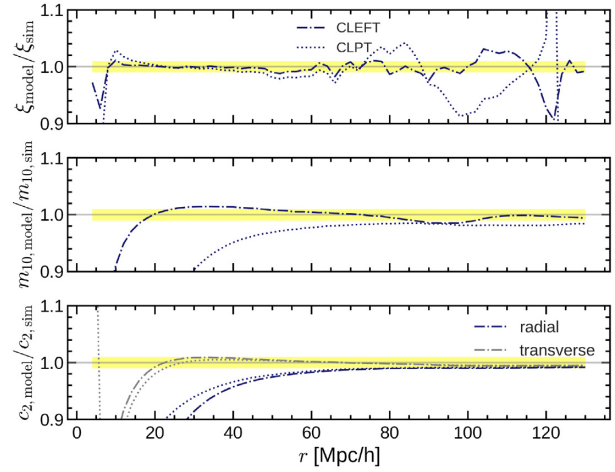


Figure C1. Detailed comparison of the different predictions for the GSM ingredients made by CLPT and CLEFT. The top panel shows the ratio of the predicted two-point correlation function to the measurement in the simulation, for both CLPT (dotted) and CLEFT (dot-dashed line). The solid yellow bands mark the one per cent agreement. The middle and bottom panels show the same comparison for the mean radial velocity, and the second-order radial and transverse moments.

CLPT and CLEFT, however, since $c_{20} = m_{20} - m_{10}^2$, the variance of the radial component is influenced by the predictions of the mean. Coincidentally, the error made by CLPT in the mean improves the agreement with the variance of the radial component (dotted blue line in the lowest panel).

Finally, we show the redshift space monopole and quadrupole in Fig. C2, obtained by combining these predictions with the GSM. The CLEFT predictions of the monopole and quadrupole are more accurate than those from CLPT, mainly due to the increased accuracy in estimating the mean pairwise velocity, which is consistent with our findings in Section 5. As shown in Section 3.4, on scales smaller than 30 $h^{-1}\text{Mpc}$ it is necessary to include higher order moments to further improve the accuracy of the predictions. A more detailed comparison of these different models applied to mock catalogues that mimic actual data at different redshifts and different halo mass ranges will be the subject of future work.

Table C1. Perturbation theory parameters for both CLEFT and CLPT. Note that b_1 , b_2 , and b_s are obtained by expanding the bias function in Lagrangian space. We show the maximum likelihood estimate and errors representing 1σ deviations in the posterior distribution of the given parameter.

	b_1	b_2	b_s	α_ξ	α_v	σ_{FoG}
CLPT	0.29 ± 0.01	-1.63 ± 0.31	1.80 ± 0.38	-	-	-17.35 ± 0.15
CLEFT	0.30 ± 0.02	-1.69 ± 0.26	2.16 ± 0.37	-39.58 ± 16.32	90.30 ± 73.68	-17.45 ± 0.28

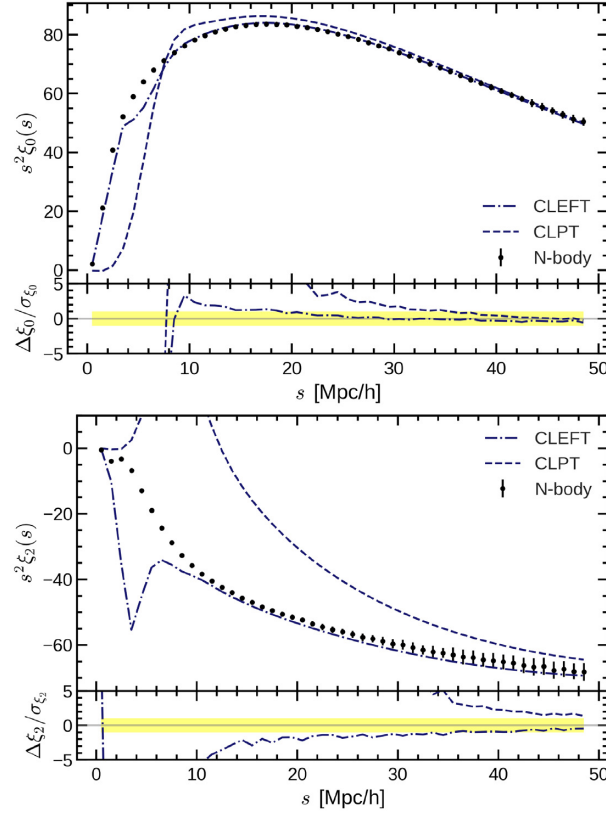


Figure C2. Comparison of the GSM predictions for the redshift space monopole and quadrupole, using the real space ingredients predicted by CLPT and CLEFT. The residuals are plotted as the difference between the model and the simulation in units of the variance calculated across the different independent simulations. The yellow bands show the 1σ deviation.

This paper has been typeset from a \LaTeX file prepared by the author.